THE VALIDITY OF THE HERITABILITY CONCEPT IN QUANTITATIVE GENETICS

By

ASKO MÄKI-TANILA

Doctor of Philosophy

University of Edinburgh

1982

To Kukola

ABSTRACT

A substantial positive curvature was discovered in offspring-parent regression for a bristle number in a laboratory population of Drosophila melanogaster. This initiated a theoretical investigation of the restrictions on the classical prediction equation in quantitative genetics. The main interest was in models of genetic and environmental variation which lead to asymmetrical responses in opposite directions in the first generation of selection, that is to non-linearity in offspring-parent regression. Non-linearity was assessed by fitting a quadratic regression in an infinite random mating population.

A genetic model with a small number of loci, each with an arbitrary number of alleles, was used. The effects of dominance, multiplicative interaction and unequality of loci were studied. When there are no environmental deviations, apart from the case of complete dominance, single parent and mid-parent regression were found to show similar curvature, so that in general the offspring-parent regression between genotypic values has largest departures from linearity when there are rare, almost completely recessive alleles segregating at equal loci or in an analogous way directional recessivity and low averaged gene frequency over unequal loci. When the number of alleles is increased non-linearity decreases. For the number of loci making equal contributions to the variation the amount of non-linearity is roughly proportional to 1 / number of loci.

To study the effect of an additive independent environmental deviations various distributions for them were used. It was shown that when the offspring-parent regression between genotypic values

is linear and H is the ratio of genotypic to phenotypic variance, the regression of offspring on parental phenotype is linear only if the skewness of the environmental distribution is a proportion $\sqrt{H^2/(1-H^2)}$ of that of the genotypic distribution. The more the skewnesses depart from this equality or smaller H^2 is, the larger the departures from linearity are. The genotypic non-linearity shows up only if H^2 is very large. When the environmental deviations have a normal distribution, the largest departures from linearity are expected when there are rare and completely recessive alleles segregating at loci with large contributions. Models with dependence between genotypic and environmental distributions were also studied.

Multiplicative interaction as such was shown to make only small contributions to non-linearity, its effect being more substantial when there is a locus with a very large effect acting in a genetic background due to a very large number of loci with small effects.

The use of Abplanalp's linear heritability estimates in checking the asymmetry of response were examined fitting a quadratic regression between sibs. Non-linearity in half-sib regression was found to be the same as in the regression of offspring on single parent. Dominance and common environmental effects were shown to cause biases in full-sib estimates. The effects of linkage on sib on sib regression were discussed when there is a large number of multiplicative loci contributing to the variation.

CONTENTS

1. INTRODUCTION

The difference between evolutionary theory of population genetics, based on the study of well-identified loci, and the statistical theory of quantitative inheritance, directed to the analysis of phenotypically measured characters, is widely recognized. The former has become largely concerned about expectations for the change of gene and chromosome frequency under drift and various modes of selection, the effects of which are usually difficult to measure in natural populations. The latter traditionally makes (linear) assumptions about the interactions of the various forces which might affect a phenotype, and in particular about the contributions of genetic factors. These assumptions, which originate from Fisher (1918) and Wright (1921), are also extremely difficult to justify empirically. The phenotype is taken to be a sum of individual gene effects, plus an environmental effect, using the term environment in its widest sense. The variance of such a sum can be computed from the variances and covariances of its individual terms, which in turn depend on gene effects, gene frequencies and pairwise linkage disequilibria.

Most of the theoretical quantitative genetics has its interest in consequences of controlled modification of existing phenotypic variation, in either plant or animal populations. Due to differences in immediate applications and in initial preferences, two schools of thoughts exist, which from time to time forget their genuine divergences and plunge into most futile confrontations (e.g. Kempthorne, 1977; Jinks, 1979).

The development of the part of quantitative theory relevant to plant breeding, associated mainly with names of Mather and Jinks, relies heavily on the analysis of genetic differences between homozygous strains and of the variation in segregating generations derived from crosses between the strains (Mather and Jinks, 1971). The school claims to have developed analyses which cover almost every conceivable situation and which are able to detect and often estimate the contribution of all known sources of variation (Jinks, 1979). If true we have a mechanistic model which, given a controlled breeding program, allows us to recapture the past or predict the future without being much in error. However, the range of organisms to which such analyses can be applied is severely restricted by the necessity of having highly inbred lines as starting point. Since the school's primary interest is in the properties of genes controlling quantitative variation, it is clearly appropriate to use homozygous lines. However, only random breeding population such parameters can, at best, refer to will be either a population derived from the inbred lines or a population from which the inbred lines have been derived.

The classical theory of quantitative genetics, as applied to animal breeding, is due primarily to Sewall Wright and was very much extended by Lush (e.g. Robertson, 1969). It gives a statistical description, in terms of variance components, of the genetic and environmental variation affecting a particular measurement in a random breeding population as it is at the moment and allows some short-term predictions of the response to selection and, with less adequacy, of the effect of inbreeding. In essence, the introduction of only one new concept, that of the additive genetic variance and

its related parameter, heritability, provides a coherent framework into which different observations that can be made on the population, such as the effect of selection, or the similarity between relatives, can be fitted. This leads to the classical prediction equation of quantitative genetics, namely that

$$\Delta G = h^2 \Delta P,$$

where ΔG is the expected genetic change produced by selection differen ial ΔP, the superiority of selected parents above the mean of the population from which they were chosen, in a trait with heritability h^2. The equation can be also written, as

$$\Delta G = i\ h^2 \sigma_P, \qquad (1)$$

where σ_P^2 is the phenotypic variance and i the standardised selection differential, i.e. $\Delta P/\sigma_P$. These equations imply that selection in opposite directions yields symmetrical responses and that the relationship between the response and selection intensity is linear, that is to say heritability or the additive variance is constant over the phenotypic range. It is the validity of these arguments we shall discuss and investigate in more detail.

Early Asymmetry in Selection Experiments

Asymmetrical responses are commonly observed in selection experiments, especially when selection has lasted for several generations. We consider the main causes to be the following (Falconer, 1981):

i. Random drift produces changes of gene frequencies which are cumulative over generations. Even a given gene content does not necessarily yield a predictable outcome, because during selection genes can be lost through sampling. Thus a selected line may become fixed for a particular allele, even though a better one had originally been available, because the better allele was accidentally lost in the process. This accidental loss will be less likely as population size, frequency of the favourable allele, magnitude of allelic effect increases (Robertson, 1960). Hill's studies of the sampling variance of the selection response (e.g. Hill, 1971) show that drift can be one of the most important causes of observed asymmetry in small unreplicated experiments (cf. Falconer, 1973).

ii. If we call allele frequencies, at which heritability for a given degree of dominance is at its maximum, symmetrical frequencies, there will be asymmetrical responses to upward and downward selection if frequencies are on the average above or below these values.

iii. If a character is a component of fitness, selection towards increased fitness may give a slower response, since the loci affecting such a character are likely to show dominance (Robertson, 1955),the trait being therefore subject to inbreeding depression , and also, the allele frequencies at these loci are likely to be high.

iv. If the variance is dependent of the mean, i.e. there is scale effect, asymmetry follows.

v. The trait can have a maternal component, and if this is different for high and low values, responses in opposite directions will be asymmetrical.

Most of the effects mentioned above usually result in asymmetrical responses only in long-term experiments. There is, however, some evidence on either asymmetry immediately in the early few generations of selection or a non-linear relationship between the response and selection differential in the first generation, in other words, the offspring-parent regression may be better approximated by a non-linear rather than a linear function of parental values.

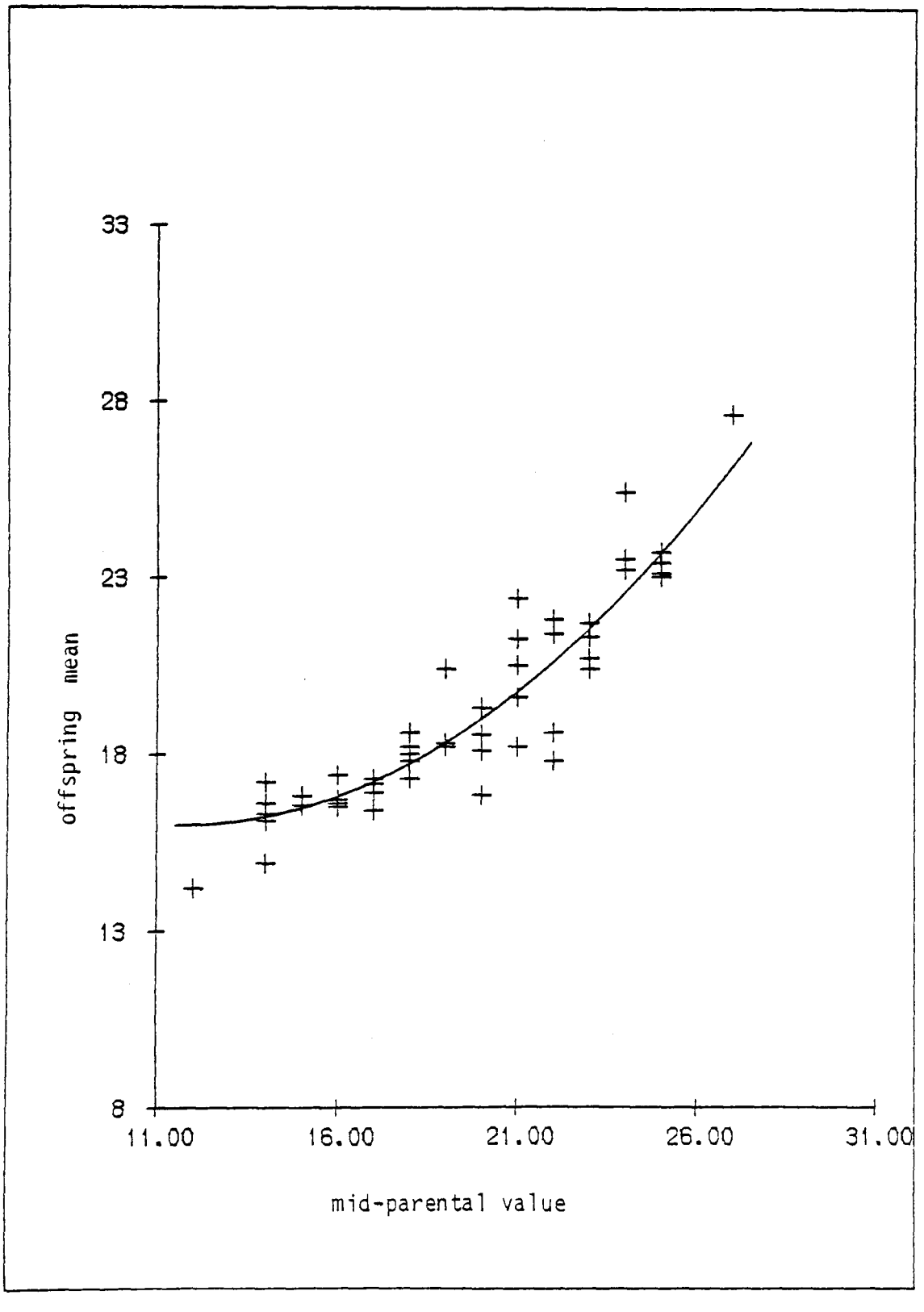
Some extensive studies have been made in dairy cattle where a large number of observations is usually available allowing for detection of small differences in calculated parameters. Beardsley, Bratton and Salisbury (1950) while studying milk fat production found evidence for lower heritability as deviations from the mean increased, however the curvilinearity of daughter-dam regression was not significant. Bradford and Van Vleck (1963) and Van Vleck and Hart (1965) showed that for a single generation response a linear relationship was adequate even with very large selection differentials, irrespective of the direction of selection. There have been many analyses of the milk yield data, in which herds have been classified and divided into two or three classes according to the average level of production and estimates of heritability have been made for each level. Generally higher heritabilities have been found for higher level of production (see Petersen, 1972, for a review). From this we can tentatively conclude, although the traits are different, that since Beardsley et al. (1950) had a small number of records and did not base their calculations on the deviations from the average herd-mates the curvilinearity may have been caused by the environmental effects.

Nishida (1972) reported on a strong non-linearity of offspring-parent regression for body weight in mouse. His results were very inconsistent in the sense that heritability varied from 0.13 to 0.62 over age classes, apparently with very large standard errors, and the regression on mid-parental value was in some cases linear, in others quadratic, with either convexity or concavity, and in some cases cubic. Meyer and Enfield (1975) carried out a large-scale experiment in Tribolium to examine the effect of selection intensity, both upward and downward, on the magnitude of realized heritability after one generation, when selection was practised in both sexes. The character studied was a 21-day pupa weight. Within the direction of selection heritability estimates were fairly constant, but a high degree of asymmetry was found for the response to two-way selection, the realized heritabilities being much larger downwards than upwards.

Experiment in Drosophila

A simple experiment was designed as a further test of the asymmetry in a single generation response reported earlier by Robertson (1977) in this laboratory. The source population was the Dahomey cage population of Drosophila melanogaster which had been kept in the laboratory for several years prior to this experiment. The trait scored was the sum of the number of sternopleural bristles on the right- and left-hand side, in both males and females. The cage was sampled by inserting culture bottles. Two hundred flies of each sex emering from these bottles were scored, the mean being 18.55 and variance 5.86. The distribution was skewed to the right, but deviations from normality were not significant. 45 pair matings were made assortatively, with the correlation between mates being

Figure 1
Quadratic regression of offspring mean on mid-parent value for the number of sternopleural bristles in Dahomey population of Drosophila melanogaster. Each cross represents the mean of a family comprising five female and five male progeny.



virtually one, and the progeny were reared in vials. Five male and five female progeny of each sex were sampled at random and scored for the bristle number. When regressing the progeny mean on mid-parental value, the slope of the linear regression was found to be 0.71 ± 0.05, and the equation for the quadratic regression was $y = 21.8 - 1.00x + 0.043x^2$, where y is the predicted progeny value and x the mid-parental score. The deviations from linearity were significant ($F = 9.63 > F_{5\%} = 4.08$). In Figure 1 the progeny means are plotted against mid-parental value and also the regression curve is shown.

Response at Individual Loci

With respect to loci affecting a quantitative trait , response to selection involves two processes, a change in allele frequencies following selection, and a consequent change in population mean. So, in deriving the equation (1) in terms of effects from individual loci we divide the prediction in two steps (Robertson, 1977). We calculate the expected changes of allele frequencies at the relevant loci in the selected parents produced by a given selection differential, and then we predict the change in the population mean produced by the new allele frequencies.

Change in Allele Frequency

There are different ways to derive the change in allele frequency at a single locus. Most simply it is obtained by a linear regression argument (Robertson, 1963), but a more generalized derivation can be done using probabilities of selection (Kimura,

1958), the most recent discussion on this being by Kimura and Crow (1978).

Let us suppose that, except for a small displacement in mean, different genotypes at a locus have the same phenotypic distribution. This can be because there are many loci or large environmental effects contributing to the variation, or both. Truncation selection is practised, so that individuals with values above a certain culling level are chosen and the rest is rejected, the proportion saved being p. For an individual with a genotypic value Y, the probability that it is selected, f(Y), can be written by the Taylor's expansion about the population mean, M, as

$$f(Y) = f(M) + (Y-M)f'(M) + (Y-M)^2 f''(M)/2 + \ldots$$

The relative probability of selection, or the fitness, of a genotype is then $f(Y)/p$. If the differences in the genotypic values are very small, so that terms in $(Y-M)^2$ and higher order can be ignored then

$$f(Y)/p = 1 + (Y-M)Z/p,$$

where $Z = f'(M)$, the ordinate of the density function at the truncation point. The change of a particular allele frequency is obtained by summing over genotypes and weighting the fitnesses by the number of the allele carried. Let us assume that there is a locus with two alleles, B and b, which contribute a ($a > 0$) and 0, respectively, to the measurement of a trait, without any dominance, and denote the frequency of B by q, and the change in

it by Δq. We find then

$$\Delta q = q(1-q)a\ Z/p.$$

In other words the change in gene frequency is proportional to its effect multiplied by the ratio of the ordinate at the truncation point to the proportion selected. If the measurements are approximately normally distributed, $Z/p\sigma_P = i$ and we have a further simplification, namely

$$\Delta q = i(a/\sigma_P)q(1-q). \qquad (2)$$

This result was first derived by Haldane (1931). The quantity a/σ_P in equation (2) is half the difference of value between the two homozygotes, expressed in terms of the phenotypic standard deviation, it can be referred to as the standardised effect of the locus (Falconer, 1981). Hence the regression of the allele frequency on the phenotypic value is linear for small a/σ_P (Robertson, 1963), whereas for non-normal distributions this regression is generally not linear (Cockerham and Burrows, 1980).

Assuming normality of phenotypic distribution, Latter (1965) has studied the consequences when there are loci with large effects contributing to the variation, by including the second term of the Taylor's expansion. Then $f''(M) = xz/\sigma_P$, where x is the truncation point and z the ordinate of the standardised normal distribution. Hence the fitness of a genotype with value Y is

$$f(Y)/p(1+ix(a/\sigma_P)^2q(1-q)) = 1 + i(Y-M)/\sigma_P + ix(Y-M)^2/2\sigma_P^2,$$

and the change in gene frequency is

$$\Delta q = i(a/\sigma_p)q(1-q)[1 + x(a/\sigma_p)(1-2q)/2]. \qquad (3)$$

Assuming that (3) is the realized change in gene frequency, Latter showed that compared with (2) for loci with large effects relative to standard deviation, the second-order terms are important in predicting only with intense selection and extreme gene frequencies, (2) and (3) being equal if half is selected or $q = 0.5$.

We allow for dominance and assume that the heterozygote Bb deviates by da from the mean of the two homozygotes, where d, degree of dominance, is interpreted as follows:

$d = 0$, no dominance,

$0 < |d| < 1$, partial dominance,

$d = \pm 1$, complete dominance.

Heterozygote superiority or inferiority (i.e. $d > 1$ or $d < -1$) is ignored.

Then for a normal phenotypic distribution the change in gene frequency for small a/σ_p is

$$\Delta q = i(\alpha/\sigma_p)q(1-q), \qquad (4)$$

where $\alpha = (1+d(1-2q))a$, the average effect of a gene substitution (e.g. Falconer, 1981) at this locus. This was first derived using probabilities of selection by Kojima (1961), and can be obtained

also by the regression method (Robertson, 1977). When the standardised effects are large, we have by taking also the second-order terms in the Taylor's expansion (cf. Silvela, 1980), as

$$\Delta q = i(\alpha/\sigma_p)q(1-q)\left[1 + x(\alpha/\sigma_p)\ (1-2q)/2\right] - 2ix(\alpha/\sigma_p)^2 daq^2(1-q)^2. \quad (5)$$

Beside the change in gene frequencies selection causes correlations in gene frequencies between loci, i.e. linkage equilibrium (Bulmer, 1971). But since this affects heritability only from the second generation onwards, we shall in concentrating on the properties of the response in the first generation of selection, ignore these effects here.

Change in Mean

The relationship between population mean and allele frequencies is completely deterministic, i.e. the population mean is an exact function of allele frequencies. The problem of assessing the expected gain can be divided here into two major parts, whether the selection criterion in male and female parents is different or the same.

Let us write for the gene frequency in males and females as q_m and q_f, respectively, and for the mean of the contributions from a single locus as $\underline{M}(q_m, q_f)$. Suppose that in the base population $q_m = q_f = q$ and that the selection has produced changes Δq_m and Δq_f, where Δq_m does not necessarily equal Δq_f. Then if we write for the rth partial differential operator, as

$$\underline{M}^{(r)} = (\Delta q_m\, \partial/\partial q_m + \Delta q_f\, \partial/\partial q_f)^r \underline{M},$$

we have for the progeny mean after selection by the Taylor's expansion about (q,q), as

$$\underline{M}(q+\Delta q_m, q+\Delta q_f) = \underline{M}(q,q) + \underline{M}^{(1)}(q,q) + \underline{M}^{(2)}(q,q) + \ldots$$

The same approach can be easily extended to cater for multiple alleles and for any level of ploidy (Li, 1976). Since under random mating for the progeny $\underline{M} = (q_m + q_f)a + (q_m + q_f - 2q_m q_f)da$, the first and second derivatives of $\underline{M}$ with respect to q_m and q_f are

$$\partial \underline{M}/\partial q_m = a + (1-2q_f)da, \qquad \partial \underline{M}/\partial q_f = a + (1-2q_m)da,$$

$$\partial^2 \underline{M}/\partial q_m \partial q_f = \partial^2 \underline{M}/\partial q_f \partial q_m = -2da$$

Since we have diploidy there are no interactions including three or more alleles within a locus, thus derivatives of higher than second order all equal zero. Also, $\partial^2 \underline{M}/\partial q_m^2 = \partial^2 \underline{M}/\partial q_f^2 = 0$. Hence if the selection is the same in both sexes, the change or the response in performance due to the change in gene frequency at one locus is

$$\Delta \underline{M} = 2(1+(1-2q)d)a\,\Delta q - 2da(\Delta q)^2$$

$$= 2\alpha\Delta q - 2da(\Delta q)\,, \qquad (6)$$

and if there is selection only in one sex,

$$\Delta \underline{M} = \alpha\,\Delta q. \qquad (7)$$

The same results have been obtained by Robertson (1977) in a

slightly more intuitive way.

If the gene effects are small relative to the phenotypic standard deviation and we can thus ignore terms in $(a/\sigma_P)^2$ or higher order, (6) becomes, from (4),

$$\Delta \underline{M} = 2i\alpha^2 q(1-q)/\sigma_P .$$

Summing over loci and noting that $2 \sum_j \alpha_j^2 q_j(1-q_j)$ is the additive genetic variance, we obtain

$$\Delta G = \sum_j \Delta \underline{M}_j = i h^2 \sigma_P ,$$

as in (2). Had the selection been practised only in one sex, the response with the same intensity of selection is half of this. Thus, with small effects, the response is linearly related to the selection intensity irrespective of gene frequency and dominance. These approximations cannot be expected to hold for genes with larger effect. By including second order terms in (a/σ_P), we have for the response due to a single locus , from (5) and (6), when selection is the same in both sexes, the expression

$$\Delta \underline{M} = 2i\alpha^2 q(1-q)/\sigma_P + ix\alpha(\alpha^2 q(1-q)(1-2q)$$

$$- 4da^2 q^2(1-q)^2)/\sigma_P^2 - 2dai^2 \alpha^2 q^2(1-q)^2/\sigma_P^2$$

and from (7), if selection is carried out only in one sex, the expression

$$\Delta\underline{M} = i\alpha^2 q(1-q)/\sigma_P + ix\alpha(\alpha^2 q(1-q)(1-2q)-4da^2q^2(1-q)^2)/2\sigma_P^2.$$

For example, if the variance is due to equal loci with complete dominance and $q = 0.9$ and $2a/\sigma_P = 1$, the response upward, if selection is practised in one sex only, is going to be overestimated from the conventional prediction by 31.4% and by 20.6%, when the proportion selected is 10% and 20%, respectively. However, even if the assumption about the normality was true, which is hard to validate with extreme gene frequencies and complete dominance, this approximation cannot be expected to be of much value (cf. Sørensen, 1980), and therefore the results on asymmetry must be treated with some caution.

The generalization of this approach to deal with epistasis is not straightforward. When there are allelic interactions also between loci, the mean is a function of linkage equilibrium between loci, as well as of allele frequencies at individual loci. Thus, instead of changes in allele frequencies we must rather consider the changes in gamete frequencies. Another difficulty arises, namely in the presence of epistasis the distributions for genotypes are different, whereas in calculating the probability of selection it is required that these distributions differ only in means, not in shapes. In general, as noticed by A. Robertson (personal communication) and by Cockerham and Burrows (1980), even the more general formula for changes in gene frequencies, derived by Kimura and Crow (1978), can be used only when there is a large number of additive loci in linkage equilibrium causing the variation, with the environmental effects being independent of genotypic values, although an allowance for a non-normal environmental distribution can be made. Hence,

when we want to consider models with only a small number of loci, epistasis, or any kind of dependence between genetic and environmental distributions, another method is required.

Regression of Genotype on Phenotype

The problem of asymmetry can also be viewed for all loci together. Assuming that the genotypic values between offspring and parent are linearly related, asymmetry can still arise, if the regression of genotypic value on phenotypic value (in short, genotype on phenotype) amongst parents is not linear.

Extending Lindleys's (1947) theory on the linearity of regression when variables are subject to error, Curnow (1960) has, in essence, shown that given a normal distribution of environmental deviations the genotypic value is linearly related to the phenotypic value only if the genotypic distribution is normal, and also that the degree of non-linearity depends on the ratio of the environmental to the phenotypic variance and on the amount of non-normality of the genotypic distribution.

A related result was empirically found by Nishida and Abe (1974). They showed that the regression of genotype on phenotype is approximately linear, if the skewnesses of genotypic and environmental distributions are equal, and non-linear if they are not, the deviations from linearity being larger, the larger the difference between the skewnesses is. Furthermore, when the genotypic skewness is larger than the environmental one, the regression will be convex, and when less concave. Comparing linear

and quadratic regression of genotype on phenotype Robertson (1977) showed that this result can be treated as a special case of Lindley's (1947) theory.

The purpose of this work is to study by means of offspring-parent regression the type and amount of asymmetry in response to selection in the first generation. The non-linearity of offspring-parent regression, as such, has only briefly been mentioned in discussions of the theory of quantitative genetics (e.g. Kempthorne, 1960). It is only very recently that a more detailed study was carried out by Robertson (1977), which was reviewed with some amendments by Bulmer (1980). The present work is aimed to develop Robertson's treatment on the subject by investigating further, how, to what extent and under what kind of models of genetic and environmental variation the offspring-parent regression is likely to depart from linearity. We shall also consider ways to detect asymmetry when observations are available only on a single generation of a pedigreed population.

2. NON-LINEARITY OF OFFSPRING-PARENT REGRESSION

In this section the ways of finding the regression function of offspring on parent, especially the one between genotypic values, are discussed. The methodology proposed will form a basis for the study of different models which follows.

Some Exact Regression Theory

Consider a large random mating diploid population and a quantitative character determined by n autosomal loci each with two alleles, B and b. Suppose that these alleles contribute to the measurement of this trait a and 0, respectively, without any dominance. The frequency of the allele with a higher value is q at each locus. If G_i is the contribution to the genotypic value from the ith locus (G_i = 0, a, 2a for the genotypes bb, Bb, and BB, respectively), under additivity of effects over loci the genotypic value, G, is $\sum_{i=1}^{n} G_i$, and the mean genotypic value, M, is 2nqa.

Let us consider offspring from males which are mated randomly to females within the population. Since the offspring receives one of the two paternal alleles at random, and receives B or b from the female parent with probabilities q and 1-q, the expected value of the contribution of the ith locus to the offspring's genotypic value, O_i, given the contributions from all the loci in the male parent, G_{mj} (j = 0, ..., n), i.e. the regression of the former on the latters, is

$$E(O_i|G_{m1}, G_{m2}, \ldots, G_{mn}) = qa + G_{mi}/2.$$

This is because loci are assumed to be independent of each other, that is to say that there is neither epistasis nor linkage equilibrium between loci. Summing over loci gives us

$$E(O|G_m) = naq + G_m/2 = M + (G_m - M)/2.$$

Writing G_f for the genotypic value of the female parent and assuming random mating we find that the regression of offspring on both parents is

$$E(O|G_m, G_f) = G_m/2 + G_f/2,$$

from which it is obvious that the regression on the mid-parental value, $\bar{G} = (G_m + G_f)/2$, considered as a single variable, is

$$E(O|\bar{G}) = \bar{G}.$$

Thus the offspring-parent regression between genotypic values is linear under a complete additivity, for any allele frequency or number of loci.

We allow for dominance assuming that B is completely dominant over b ($d = 1$). Hence $G_i = 2a$, if the genotype at the ith locus is BB or Bb, and $G_i = 0$ if the genotype is bb, and $M = 2naq(2-q)$. We shall first find the regression of offspring on one parent. In the male parent the number of loci, whose contribution to the total value is 2a, is $G_m/2a$, and there are $n - G_m/2a$ of

those with 0. The two cases for the conditional probability that $O_i = 2a$ for a given G_{mi} are

$$\text{Prob}(O_i=2a \mid G_{mi}=2a) = (1+q(1-q))/(2-q)$$

and

$$\text{Prob}(O_i=2a \mid G_{mi}=0) = q.$$

Because of the independence we have

$$E(O|G_m) = \left[(1+q(1-q))G_m/((2-q)2a) + q(n-G_m/2a)\right]2a$$

$$= M + (1-q)(G_m-M)/(2-q). \qquad (8)$$

Similarly it can be shown that if B is completely recessive,

$$E(O|G_m) = M + q(G_m-M)/(1+q).$$

Hence complete dominance of either allele does not affect linearity of genotypic regression when only one parent is considered.

We turn to the effect of dominance on the mid-parent regression and consider first the regression of offspring on both parents. Through some elegant lines of algebra we obtain a formula, first derived by Pearson (1904), as

$$E(O|G_m,G_f) = (G_m+G_f)/(2-q) - G_m G_f/2na(2-q). \qquad (9)$$

This is a hyperboloid surface tending to a plane when n is large. The product $G_m G_f$ in (9) can be written as

$$(G_m+G_f)^2/4 - (G_m-G_f)^2/4 = \bar{G}^2 - 2na^2q(2-q)(1-q)^2.$$

Hence we have an approximation for the mid-parent regression, as

$$E(O|\bar{G}) = q\ (1-q)^2a + 2\bar{G}/(2-q) - \bar{G}^2/2na(2-q),$$

clearly a non-linear function of mid-parental value. The quadratic term in the regression is seen to vanish when the number of loci becomes very large. When the allele with a larger effect is completely recessive we have

$$E(O|\bar{G}) = (M-q(1-q)a)/(1+q) + 2q\bar{G}/(1+q) + \bar{G}^2/2na(1+q). \qquad (10)$$

Hence, the mid-parent regression has negative curvature when the allele with larger effect is dominant, and upwards when recessive. These offspring-parent regressions imply that, in a case of complete dominance, we will have asymmetry of the response in the first generation when selection has been practised in both sexes, whereas responses will be symmetrical if selection has been applied in one sex only.

This section dealing with the exact regression functions follows by and large Bulmer's (1980) treatment on the subject.

The exact regressions we have been dealing with so far are easy to find only if the variation is caused by equal biallelic loci showing either pure additivity or complete dominance. Also, the functions we find are not always sufficiently obvious to interpret in terms of asymmetrical responses or to make comparisons between different cases. For these reasons we approximate the exact regression function by fitting a polynomial in the independent variable. It is unlikely that the regressions we are dealing with are highly non-linear, therefore it is reasonable to suppose that the regression can be approximated to a good degree of accuracy by a quadratic curve. A polynomial of any higher order could be rather impracticable in terms of predicting asymmetrical responses and would demand very large experiments to detect.

The mean square quadratic regression between the genotypic values of offspring and one parent only can be written as

$$O = b_o + b_1 G + b_2 G^2 + e. \qquad (11)$$

If we write V_G and V_{G^2} for the genotypic variance and for the variance of squared genotypic values, respectively, the values of b_1 and b_2 which minimise $E(e^2)$ are

$$b_1 = \left[V_{G^2}\mathrm{Cov}(O,G) - \mathrm{Cov}(G,G^2)\mathrm{Cov}(O,G^2)\right] / T,$$

$$b_2 = \left[V_G \mathrm{Cov}(O,G^2) - \mathrm{Cov}(G,G^2)\mathrm{Cov}(O,G)\right] / T,$$

where $T = V_G V_{G^2} - \mathrm{Cov}(G,G^2)^2$. If we use small letters for the deviations from the overall means, i.e. $g = G - M$ and $o = O - M$,

and write $b_1' = b_1 + 2b_2 M$, (11) is found to be

$$o = -b_2\sigma_G^2 + b_1' g + b_2 g^2 + e, \qquad (12)$$

where σ_G is the genotypic standard deviation. Correcting for the scale, (12) is in a standardised form, as

$$o/\sigma_G = -b_2\sigma_G + b_1'(g/\sigma_G) + b_2\sigma_G(g/\sigma_G)^2 + e.$$

Writing μ_{3G} and μ_{4G} for the third and fourth moment of the genotypic distribution, respectively, and noting that

$$\mathrm{Cov}(G,G^2) = \mu_{3G} + 2MV_G,$$

$$\mathrm{Cov}(O,G^2) = E(og^2) + 2M\,E(og),$$

$$V_{G^2} = V(g^2) + 4M(\mu_{3G} + M\,V_G),$$

$$V(g^2) = \mu_{4G} - V_G^2,$$

the coefficients are found to be

$$b_1' = \left[(\mu_{4G} - V_G^2)E(og) - \mu_{3G}E(og^2)\right]/T,$$

$$b_2 = \left[V_G E(og^2) - \mu_{3G}E(og)\right]/T, \qquad (13)$$

where $T = (\mu_{4G} - V_G^2)V_G - \mu_{3G}^2$, or

$$b_1' = \left((\gamma_{2G}+2)h^2/2 - \gamma_{1G}E(og^2)/\sigma_G^3\right)/T,$$

(14)

$$b_2\sigma_G = (E(og^2)/\sigma_G^3 - \gamma_{1G}h^2/2)\ /T,$$

where $T = \gamma_{2G} + 2 - \gamma_{1G}^2$, and $\gamma_{1G} = \mu_{3G}/\sigma_G^3$ and $\gamma_{2G} = \mu_{4G}/\sigma_G^4 - 3$ are the skewness and kurtosis of the genotypic distribution, respectively.

Let us write r for the linear correlation between offspring and parent, and R for the multiple correlation of O with G and G^2. R is defined simply as a correlation coefficient between offspring and its quadratic regression on the parent, its square thus providing us the proportion of the variance in the offspring attributable to the quadratic regression. On the assumption that the heritability is not zero, we use as a measure of non-linearity the ratio of the variance removed by fitting G and G^2 to the variance removed by fitting only the linear term, i.e. R^2/r^2. To make the measure more explicit we subtract unity from the ratio and give it the sign of the quadratic coefficient. The measure of non-linearity, denoted by S_b, is then $\text{sign}(b_2)*(R^2/r^2-1)$. Hence S_b is zero, if the regression is linear and negative when the regression curve is concave (negatively curved) and positive when the curve is convex (positively curved).

Since $Cov(o, b_1'g+b_2g^2) = V(b_1'g+b_2g^2)$ or $b_1'E(og)+b_2E(og^2)$, we have

$$R = (b_1'E(og) + b_2E(og^2))/V_G,$$

and

$$R^2/r^2 = (b_1'E(og) + b_2E(og^2))V_G/E(og)^2 .$$

We find, from (13), after simplifying that

$$R^2/r^2-1 = (V_GE(og^2) - \mu_{3G}E(og))^2/((\mu_{4G}-V_G^2)V_G-\mu_{3G}^2)E(og)^2,$$

which after dividing both members of the ratio by V_G^5 can be written, as

$$R^2/r^2-1 = 4(E(og^2)/\sigma_G^3 - \gamma_{1G}h^2/2)^2/(\gamma_{2G} + 2 - \gamma_{1G}^2)h^4 . \qquad (15)$$

Hence, comparing (14) with (15), we notice that, given the same heritability, $b_2\sigma_G$ can be used as a measure of non-linearity.

A question immediately arises, namely how does this measure relate to the posterior measures of the degree of asymmetry in response to selection. Suppose that for a metric character, which can be taken as having a standardised normal distribution, offspring have a quadratic regression on the parental phenotype, P (= M + p), i.e.

$$E(o|p) = \beta_o + \beta_1p + \beta_2p^2 .$$

If truncation selection upwards with a selection differential i is applied, $E(p^2)$ amongst the selected parents is $1+ix$ (e.g. Falconer, 1981), where x is the culling point, and we obtain for the response, ΔG,

$$\Delta G = \beta_o + \beta_1 i + \beta_2(1+ix).$$

Since the true relationship is here assumed to be quadratic, we can compute the realized response for given values of S_b and heritability. When the phenotypic distribution is normal, that is to say when γ_{1P} and γ_{2P} equal zero, we have, from (15), as

$$E(op^2) = h^2\sqrt{|S_b|}/2.$$

In a case of normal distribution p and p^2 are orthogonal, thus $\beta_1 = h^2/2$. The quadratic coefficient is found, from (14) to be then

$$\beta_2 = \text{sign}(S_b)*h^2\sqrt{|S_b|}/2\sqrt{2}.$$

The relative amounts by which the realized response differs from the prediction based on heritability, i.e. $2\Delta G/h^2 i - 1$, are shown in Table 1 for various values of S_b and proportion selected.

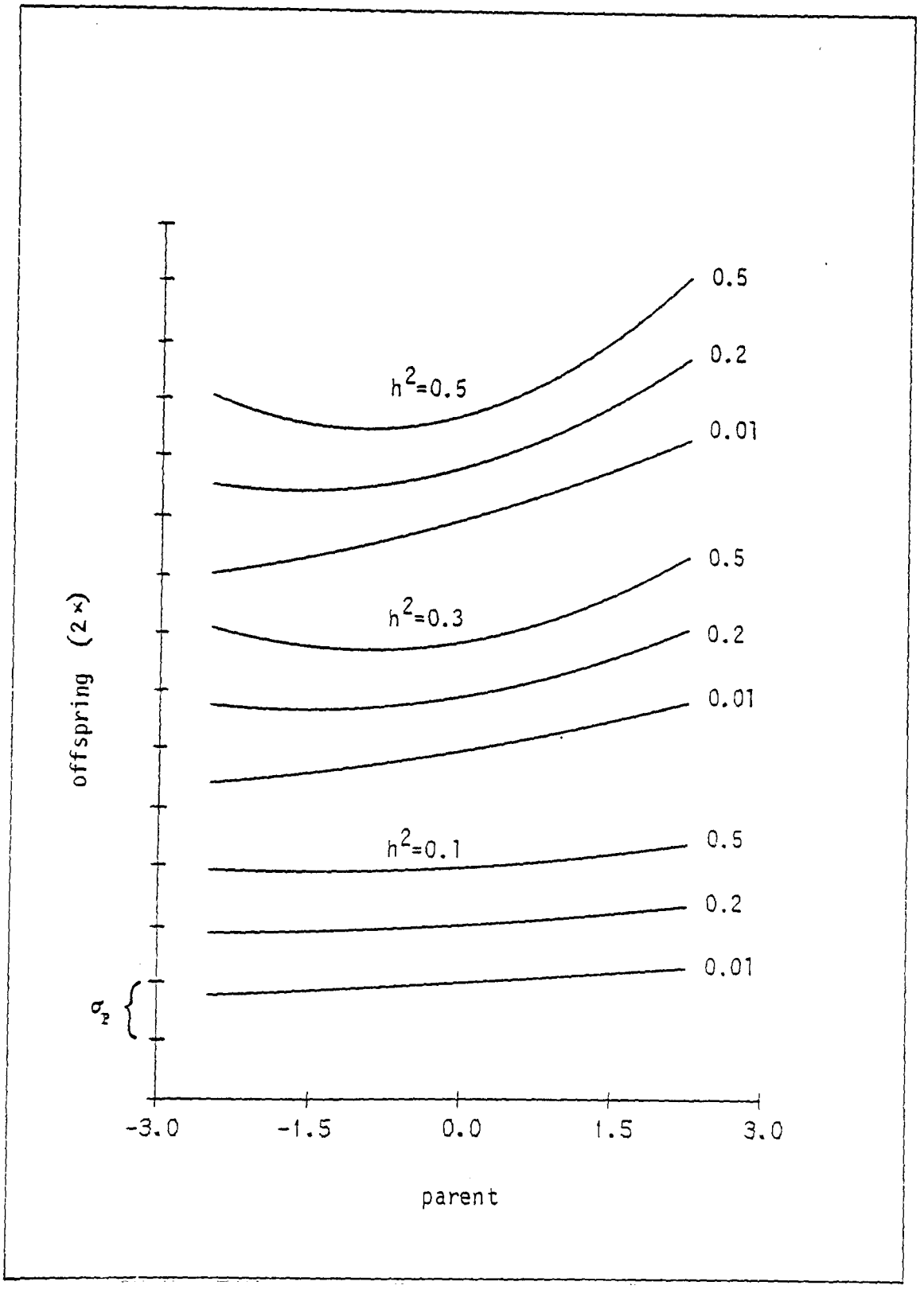
Table 1

	proportion selected	
S_b	0.20	0.10
0.01	0.06	0.09
0.05	0.13	0.20
0.10	0.19	0.29
0.20	0.27	0.45
0.50	0.42	0.64

The shapes of the regression curves corresponding to various values of S_b given in Table are shown in Figure 2 at different levels of heritability. For example, for $S_b = 0.20$ and $h^2 = 0.30$ the

Figure 2

The shapes of offspring-parent regression curves for a given non-linearity (S_b) and level of heritability when the phenotypic distribution is normal. The unit length in both axes is one phenotypic standard deviation.



realized heritability upwards would be 0.38 and downwards 0.22, when the proportion selected is 20%. Since the discrepancy between the realized and predicted responses arising from a given amount of non-linearity depends heavily on the shape of the distribution, no generalizations can be made from the normal case, which is therefore presented here only as an illustrating example.

We have been considering only single parent regression, but the same method can be used for the mid-parent regression by replacing G by $\bar{G}$ or g by $\bar{g}$.

3. GENOTYPIC REGRESSION UNDER ADDITIVE MODEL WITH DOMINANCE

Introduction

Assuming diploidy and allowing for any number of alleles per locus, the genotypic deviation can under Hardy - Weinberg equilibrium be written in a least-squares sense as $g = \alpha_i + \alpha_j + \delta_{ij}$, where $\alpha_i + \alpha_j$ is the sum of the average allelic effects making up the breeding value (or additive deviation, A) at the locus, and δ_{ij} is the dominance deviation, D, specific to the allele combination (e.g. Kempthorne, 1954). If a male with this value is mated to a large number of females chosen at random from the population then the progeny mean is, by definition, A/2. Although by definition, averaging over population, E(AD) = 0, in general $E(A^2D)$ and $E(AD^2)$ are non-zero. It can happen, for example, that individuals with large positive or negative breeding values have large dominance deviations in absolute magnitude. The agreement between the expected and actual breeding value of a son from a mating of highly selected sire and dam can be violated by large positive dominance deviations in these superior animals. In terms of bivariate moments of A and D we have for various expected values the expressions

$$V_G = E(A^2) + E(D^2),$$

$$E(og) = E(A^2)/2, \tag{16}$$

$$\mu_{3G} = E(A^3) + 3E(A^2D) + 3E(AD^2) + E(D^3),$$

$$E(og^2) = E(A^3)/2 + E(A^2D) + E(AD^2)/2.$$

The last one does not equal $E(o^2g)$ as was erroneously stated by Kempthorne (1960), instead $E(o^2g) = E(A^3)/4 + E(A^2D)/4$. The corresponding mid-parental expectations can be written, as

$$\begin{aligned} V_{\bar{G}} &= V_G/2 \\ E(o\bar{g}) &= E(og) \\ \mu_{3\bar{G}} &= \mu_{3G}/4 \\ E(o\bar{g}^2) &= E(og^2)/2 + E(og_m g_f)/2. \end{aligned} \tag{17}$$

When a male of $g_m = \alpha_i + \alpha_j + \delta_{ij}$ is mated to a female of $g_f = \alpha_k + \alpha_l + \delta_{kl}$, the progeny mean is going to be

$$(\alpha_i + \alpha_j + \alpha_k + \alpha_l)/2 + (\delta_{ik} + \delta_{im} + \delta_{jk} + \delta_{jm})/4,$$

and the covariance $E(og_m g_f)$ is found to equal

$$E((\alpha_i+\alpha_j+\delta_{ij})(\alpha_k+\alpha_l+\delta_{kl})\{(\alpha_i+\alpha_j+\alpha_k+\alpha_l)/2+(\delta_{ik}+\delta_{im}+\delta_{jk}+\delta_{jm})/4\})$$

$$= E(\alpha_i\alpha_k\delta_{ik}) = E((\alpha_i+\alpha_k)^2\delta_{ik})/2 = E(A^2D)/2, \tag{18}$$

with all other expectations equaling zero (Kempthorne, 1960). Hence

$$E(o\bar{g}^2) = E(A^3)/4 + E(AD^2)/4 + 3E(A^2D)/4.$$

Beside the moments there is another, related set of constants for describing distributions, the so-called cumulants, with properties useful from the theoretical standpoint (e.g. Kendall and Stuart, 1969). A cumulant and moment of order r are denoted by K_r and μ_r, respectively. The derivation of K_r is presented in Chapter 5. Up to the third order, moments are identical with cumulants, and, e.g. $K_4 = \mu_4 - 3V^2$. For independent random variables the cumulant of a sum is a sum of the individual cumulants, and like moments, cumulants have a property that if the variate values are multiplied by a constant c, K_r is multiplied by c^r. As moments, cumulants can be generalized for the multivariate case. Using these properties we find that

$$V_{\bar{G}^2} = V_{G^2}/8 + V_G^2/4.$$

Allowing for any kind of dominance, the additive and dominance deviations for genotypes

BB	Bb	bb

are given by (e.g. Falconer, 1981)

A =	$2(1-q)\alpha$	$(1-2q)\alpha$	$-2q\alpha$
D =	$-(1-q)^2 d$	$2q(1-q)d$	$-2q^2 d$

respectively. Hence we find, in terms of allelic effect, frequency and degree of dominance, for various expected values the expressions

$$E(A^2) = 2q(1-q)\alpha^2,$$

$$E(D^2) = 4q^2(1-q)^2 d^2 a^2,$$

$$E(A^3) = 2q(1-q)(1-2q)\alpha^3,$$

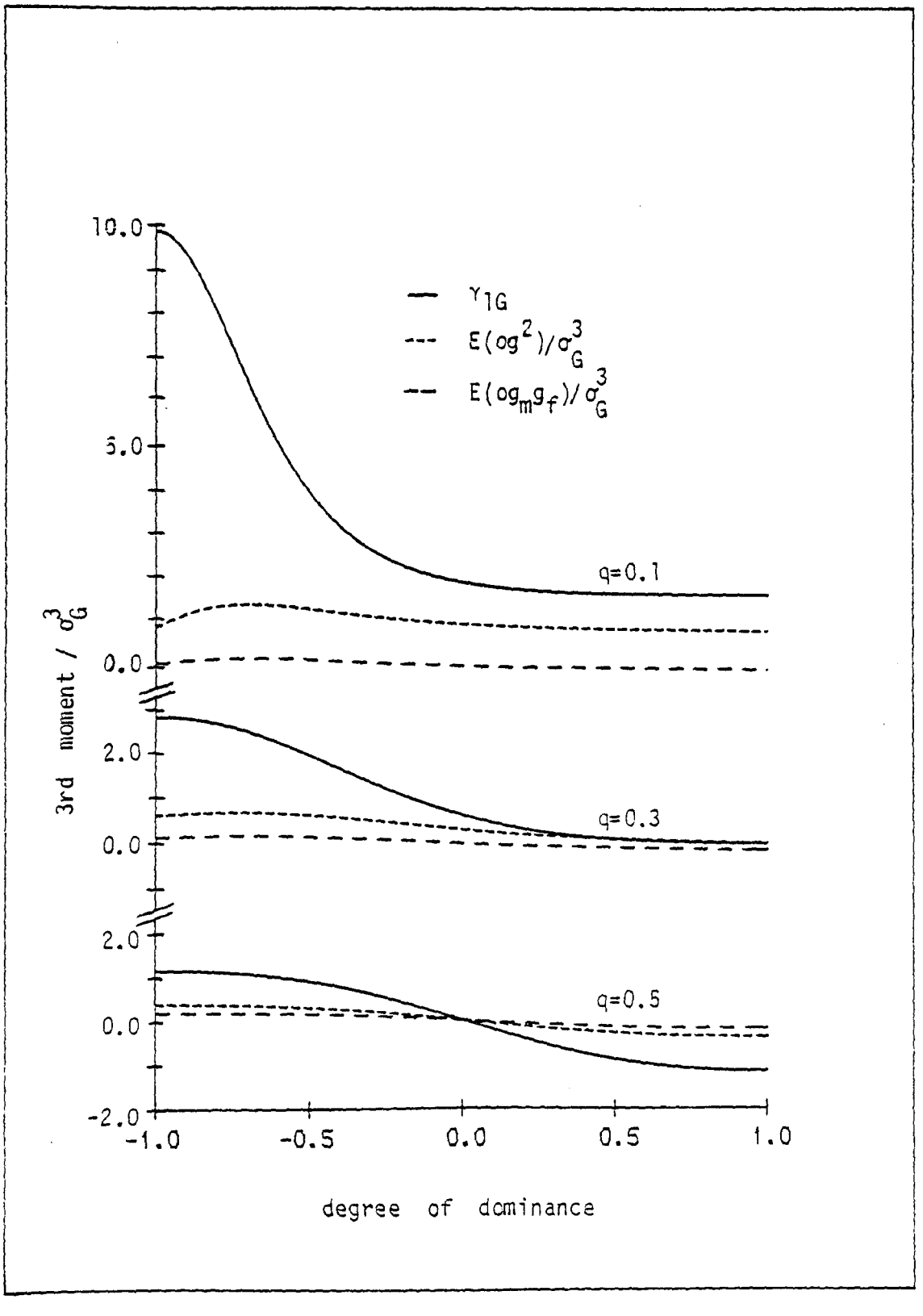
$$E(A^2 D) = -4q^2(1-q)^2\alpha^2 da, \qquad (19)$$

$$E(AD^2) = 8q^2(1-q)^2(1-2q)\,\alpha\, d^2 a^2,$$

$$E(D^3) = -8q^2(1-q)^2(1-2q)^2 d^3.$$

The third order moments are seen to be symmetrical in the sense that by interchanging the dominance relationship and the frequency between the two alleles the sign changes with the absolute values remaining the same. When $q = 1/2$, $E(A^3)$, $E(AD^2)$, and $E(D^3)$ equal zero, and when $d = 0$ so do naturally all the moments which include dominance deviation. The ways in which the gene frequency and the degree of dominance influence the magnitude of various third order moments can best be appreciated from graphical representations of the relationships derived above, in equations (16) and (19). The graphs in Figure 3 show the amounts of μ_{3G}, $E(og^2)$, and $E(og_m g_f)$, standardised by the appropriate power of σ_G, arising from a single locus with two alleles, plotted against the degree of dominance. Three cases are shown to illustrate the effect of different gene frequencies, $q = 0.1$, $q = 0.3$, and $q = 0.5$. All the parameters reach their maximum values when the plus allele is rare and recessive, and are relatively constant over the range 0.0 - 1.0 of d.

Figure 3
Magnitude of various standardised third order moments arising from a single locus with two alleles in relation to the degree of dominance at three different gene frequencies.



Models with Equal Loci

Assuming no epistasis and linkage equilibrium we have for n equal loci (subscripted by n) $E(o^r g^s)_n = nE(o^r g^s)$, where $r,s = 0, 1, 2, 3$ and $r+s \leq 3$, and

$$\mu_{4Gn} = n\mu_{4G} + 3n(n-1)V_G^2.$$

Using moments of single locus contributions, quadratic coefficient of the single parent regression can be written for n equal loci as

$$b_2 = [V_G E(og^2) - \mu_{3G} E(og)]/[(\mu_{4G} + (2n-3)V_G^2)V_G - \mu_{3G}^2] \qquad (20)$$

When the number of loci, is very large, the magnitude of this coefficient is determined by the term $2nV_G^2$ in the denominator, i.e. the regression tends to linear with a large number of loci. Using (16), the numerator of (20) can be written, as

$$E(D^2)\{E(A^3)+2E(A^2D)+E(AD^2)\}/2 - E(A^2)\{E(A^2D)+2E(AD^2)+E(D^3)\}/2.$$

When $d = 1$ this is after some simplification, from (19),

$$2q^2(1-q)^2a^2(16q(1-q)^3a^3)(1-4q+2q^2)-$$
$$2q(1-q)(2a(1-q))^2(8q^2(1-q)^2a^3)(1-4q+2q^2). \qquad (21)$$

This is equal to zero , as would be expected from (8). From the symmetry property of the third order moments mentioned above it also follows that the regression is linear when the plus allele is

completely recessive.

When the dominance is partial, non-linearity is more a rule than an exception. Figure 4 shows that the single parent regression is concave when the allele with a larger effect is recessive. Due to the symmetry in respect to q and d, the degree of non-linearity is recorded only for allele frequencies less than or equal to 0.5. The recessivity of the plus allele granted ($d < 0$), deviations from linearity are largest when the frequency is roughly $1+d$. Within this constraint the non-linearity is most substantial when the allele is almost completely recessive. When the allele frequency is intermediate (0.3 - 0.7) curvature is negligible. Corresponding statements can be made about loci with dominance in the opposite direction. In general, when the variation is caused by a small number of equal loci the non-linearity of single parent regression is most likely caused by rare, almost completely recessive alleles with the direction of dominance determining the direction of curvature. These results for the regression on only one parent have also been obtained numerically by Bulmer (1980). Our estimate of the degree of non-linearity is , in principle, the same as his, although he uses an approximation proportional to the resiprocal of the number of loci. When the allele frequency is extreme, it overestimates the rate of the approach to linearity in respect to the number of loci (Figure 5).

Non-linearity in a single parent regression can be used as an indicator of asymmetry in a two-way selection when selection is applied only in one sex. The term $E(og_m g_f)$ can be expected to cause the mid-parent regression to differ from the single parent one.

Figure 4

The relationship between non-linearity of offspring-parent regression and degree of dominance for three different gene frequencies when the variation is due to four equal loci without epistasis. Solid line refers to single parent and broken line to mid-parent regression.

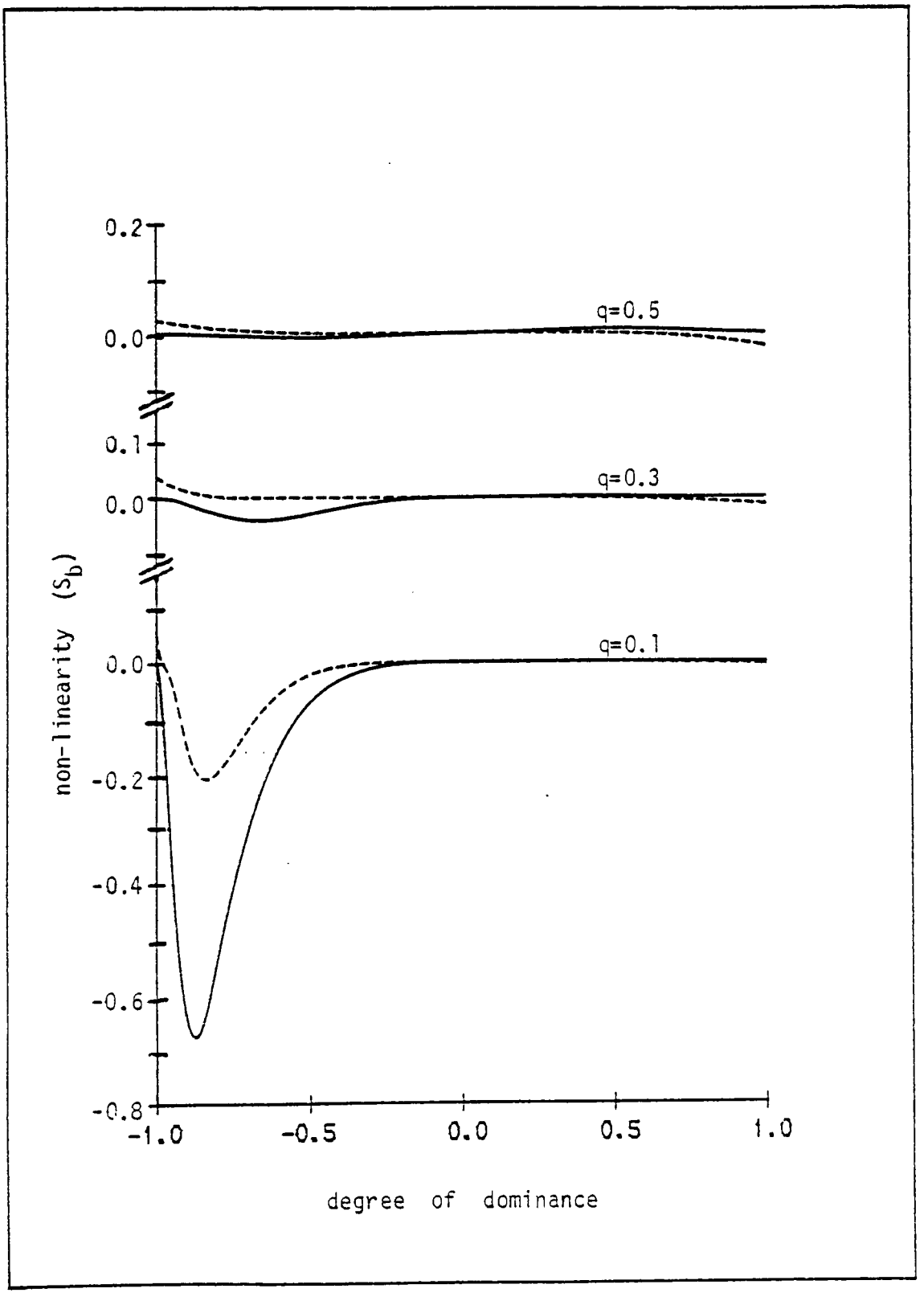
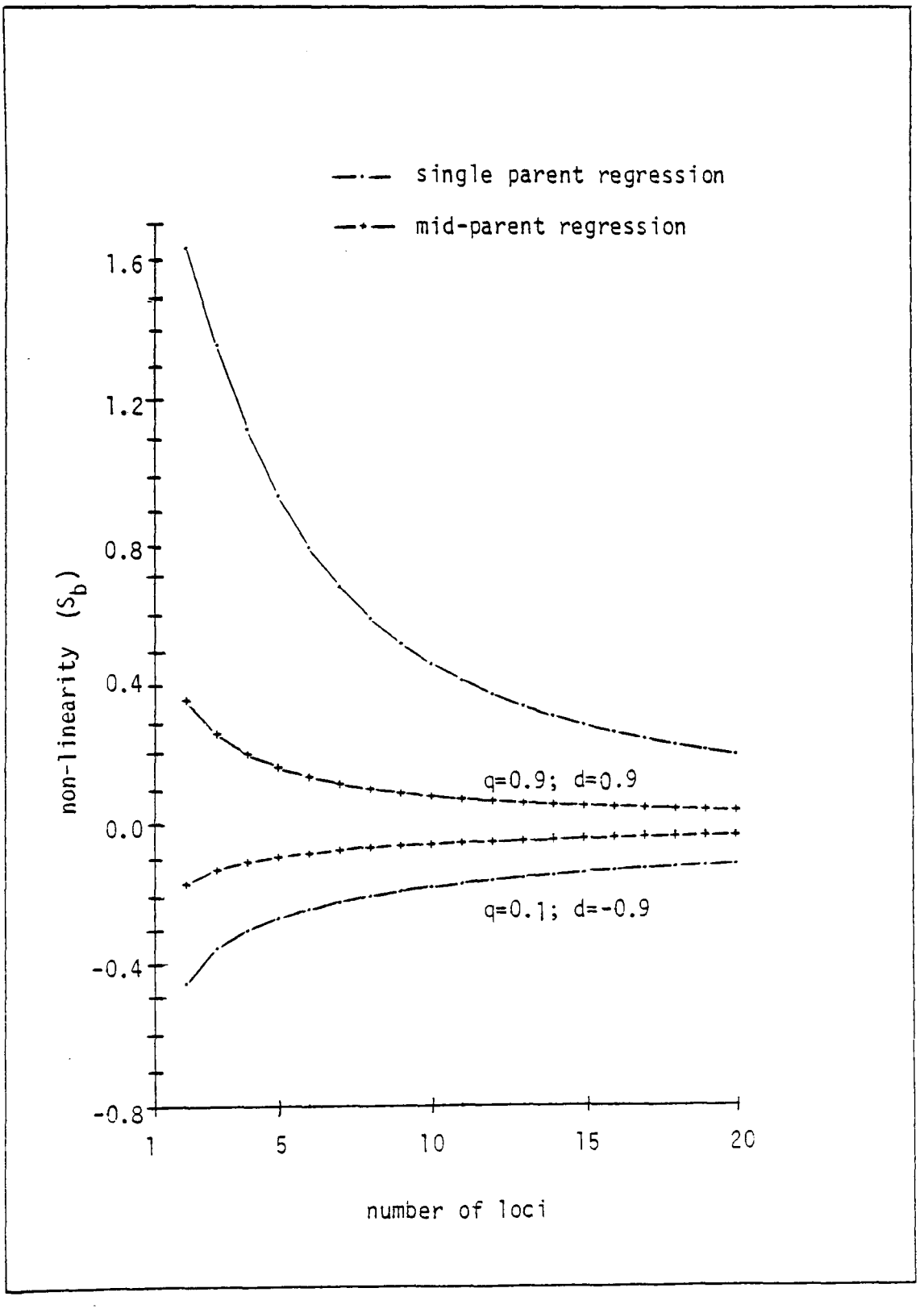


Figure 5
The relationship between genotypic non-linearity of offspring-parent regression and number of loci for a given gene frequency and degree of dominance when there is no epistasis between loci.



Hence different kind of asymmetries can be expected depending whether selection is carried out in one sex or in both sexes.

Using moments of single locus contributions, the quadratic coefficient of mid-parent regression (the bar to distinguish from the corresponding single parent parameter) is

$$b_{\bar{2}} = 4\{V_G(E(og^2)+E(og_m g_f)) - \mu_{3G}E(og)\}/\{(\mu_{4G}+(4n-3)V_G^2)V_G-\mu_{3G}^2\} \quad (22)$$

When n is large, b_2 tends to zero, despite dominance of any kind. Comparing (20) and (22), we notice that the difference between b_2 and $b_{\bar{2}}$ is proportional to the covariance between the offspring and the product of the parental values. Consequently, linearity does not follow when there is complete dominance, instead the regression is, for example, curved upwards when d = -1, the curvature being most substantial for a rare allele (Figure 4), which is in agreement with the results from the exact regressions (10). From (16) and (18), the numerator in (22) can be written as

$$E(D^2)\{E(A^3)+3E(A^2D)+E(AD^2)\} - E(A^2)\{4E(AD^2)+2E(D^3)\},$$

which implies symmetry in respect to d and q, as before. Apart from the case of complete dominance, the pattern of non-linearity is, on the whole, similar to that of single parent regression. Numerical studies show that the mid-parent regression is linear in addition to the the trivial case of pure additivity, when for

$$q = 0.02,\ 0.10,\ 0.20,\ 0.30,\ 0.40$$

$$d = -0.99, -0.98, -0.88, -0.70, -0.50,$$

respectively. Proportionality of the degree of non-linearity to the resiprocal of n serves again as a reasonable approximation except when the allele frequency is extreme (Figure 5).

Models with Unequal Loci

So far the genetic model has been simple in a sense that we have assumed allelic affects and frequencies to be equal over all loci. Let us suppose that there are n_1 loci where the frequency of an allele with a larger effect, denoted by a_1, is q_1, and the degree of dominance d_1, and n_2 loci with parameters a_2, q_2, and d_2, respectively. In the absence of epistasis and linkage disequilibrium the numerator of b_2 is

$$\left[n_1 V_{G1} + n_2 V_{G2}\right]\left[n_1 E(og^2)_1 + n_2 E(og^2)_2\right] - \left[n_1 \mu_{3G1} + n_2 \mu_{3G2}\right]\left[n_1 E(og)_1 + n_2 E(og)_2\right] \quad (23)$$

where the moments are subscripted according to the type of loci whose variation they describe. Apart from the trivial case of $d_1 = d_2 = 0$, when is this going to equal zero.

If there is complete dominance at all loci, say, $d_1 = d_2 = 1$, whilst the allelic affects, frequencies and numbers of loci, may differ between the sets of loci, it can be inferred from (21) that (23) can be written, as

$$n_1 n_2 \sum_{\substack{i=1;j=2 \\ i=2;j=1}} \{V_{Gi} E(og^2)_j - \mu_{3Gi} E(og)_j\},$$

all the rest equaling zero. Using (16) and (19), and simplifying, this becomes

$$64n_1a_1^2q_1^2(1-q_1)^2n_2a_2^2q_2^2(1-q_2)^2\sum_{\substack{i=1;j=2\\i=2;j=1}}(q_i(1-q_j)-(1-q_i)q_j)a_j(1-4q_j+2q_j^2),$$

which equals zero when $q_1 = q_2$, irrespective of differences in effects and numbers of loci.

When the dominance relationship and the frequencies are interchanged between the two alleles, the variances remain the same, but the third moments change their sign as was noticed earlier. Hence, if $d_1 = -d_2$ and $q_1 = 1-q_2$, the numerator of the quadratic coefficient is, given equal gene effects over loci, found to be

$$(n_1+n_2)(n_1-n_2)\{V_{G1}E(og^2)_1 - \mu_{3G1}E(A^2)_1/2\},$$

which equals zero when $n_1 = n_2$. Obviously this result holds also for the mid-parent regression. With these values the genotypic distribution is symmetrical.

The results for the non-linearity of offspring-parent regressions and the genotypic skewness are shown in Table 2 for the case when the variation is due to unequal loci. In constructing the table some use have been made of the symmetry properties of third order moments, e.g. the value of S_b for $q_1 = 0.9$, $d_1 = -1.0$, and $q_2 = 0.9$, $d_2 = 0.75$ equals $-S_b$ for $q_1 = 0.1$, $d_1 = -0.75$, and $q_2 = 0.1$, $d_2 = 1.0$, or S_b for $q_1 = 0.5$, $d_1 = 0$, and $q_2 = 0.9$, $d_2 = -1.0$ equals $-S_b$ for $q_1 = -0.5$, $d_1 = 0$ and $q_2 = 0.1$, $d_2 = 1.0$,

Table 2

The non-linearity ($10*S_b$) in the regression of offspring on one parent ([a]) and on mid-parent ([b]) and the skewness of genotypic distribution ([c]) when the variation is due to two loci with equal effects but unequal gene frequenies and degrees of dominance.

d_2	q_2 \ d_1 / q_1	-1.0			-0.75			0.0		
		0.1	0.5	0.9	0.1	0.5	0.9	0.1	0.5	0.9
		6.96	1.18	-1.29	5.92	1.14	-1.20	2.15	0.19	-0.64[a]
	0.1	0.00	-0.06	-0.33	-7.71	-1.13	-0.49	-0.57	-0.15	-1.09[b]
		1.36	1.20	0.00	-2.95	0.24	-0.02	-0.27	-0.07	-0.80[c]
			0.81	-0.01		0.80	0.16		0.54	0.67
-1.0	0.5		0.00	-0.15		-0.03	-0.17		-0.06	-0.15
			0.56	0.03		0.32	0.04		0.05	0.19
				-1.12			-1.12			-1.28
	0.9			-0.00			-0.00			0.00
				0.09			0.07			0.09
		5.92	1.16	-1.28	5.09	1.11	-1.18	2.01	0.18	-0.65
	0.1	-7.71	-0.05	-0.12	-5.33	-0.89	-0.20	-0.35	-0.07	-0.44
		-2.95	1.11	0.03	-2.35	0.24	0.00	-0.13	-0.03	-0.28
			0.80	-0.14		0.77	0.03		0.46	0.56
-0.75	0.5		-0.03	-0.11		-0.12	-0.15		-0.09	-0.26
			0.32	0.02		0.16	0.02		0.01	0.04
				-1.12			-1.12			-1.25
	0.9			-0.00			-0.00			-0.00
				0.07			0.05			0.04
		1.61	0.93	-0.21	1.57	0.92	0.00	1.25	0.54	0.71
	0.1	-0.07	0.00	0.00	-0.05	-0.02	0.00	0.00	-0.00	-0.00
		-0.23	0.12	0.01	-0.18	0.02	0.00	-0.04	-0.01	-0.04
		-0.86	0.11		-0.86	0.00		-0.56	-0.46	
0.75	0.5	0.00	-0.00		0.05	-0.00		0.26	0.09	
		-0.67	0.01		-0.42	-0.00		-0.04	-0.01	
		0.17			-0.00			0.65		
	0.9	-0.60			0.00			0.44		
		-0.76			-0.00			0.28		
		1.58	0.95	0.00	1.55	0.94	0.21	1.28	0.65	0.87
	0.1	-0.07	0.00	0.00	-0.05	-0.01	-0.00	-0.00	-0.00	-0.00
		-0.30	0.06	0.00	-0.26	0.00	-0.01	-0.09	-0.03	-0.08
		-0.96	0.00		-0.95	-0.11		-0.68	-0.54	
1.0	0.5	-0.20	-0.00		-0.03	0.00		0.15	0.06	
		-1.66	-0.00		-1.19	-0.01		-0.19	-0.05	
		-0.00			-0.17			0.64		
	0.9	-0.00			0.60			1.09		
		-0.00			0.76			0.80		

and so on. In addition to the results stated above, Table 2 shows that if the variation in a trait is caused by a mixture of loci, which exhibit directional dominance, that is the dominance of alleles is over loci preponderantly in one direction (Falconer, 1981), then the largest deviations from linearity follow when, roughly speaking, the average frequency of "directionally recessive" alleles over loci is relatively low. When the loci show complete dominance in the same direction the amount of non-linearity in single-parent regression is irrespective of gene frequencies fairly small, however the mid-parent regression shows substantial curvature, especially when the recessive alleles are on the average rare. The Table shows that about the genotypic skewness it generally holds that recessivity and/or rarity of an allele with decreasing effect shifts the distribution towards negative skewness.

Multiple Alleles

Models in quantitative genetics - or in population genetics in general - often make a number of simplifying assumptions, for example, that random mating takes places, that there is no linkage, that the population size is effectively infinite, and so forth. It is difficult enough to consider the effect of relaxing one of these assumptions and quite impossible to study the consequences of relaxing them all, simultaneously. We have so far taken it granted that there are, at the most, two alleles per locus segregating in a population. Next we shall remove this simplicity and study in what kind of circumstances the offspring-parent regression is likely to deviate from linearity when in the population there are more than

two segregating alternatives per locus.

Although there has been a clear demand to include multiple alleles in the models for quantitative inheritance (e.g. Lewontin, 1977), only few studies (Latter, 1969; Latter and Novitski, 1969) have made a significant allowance for them. This is probably due to the lack of any strong evidence for their existence. Robertson's (1960) theory on selection limits suggests that if we through selection fix alleles which are rare in the initial population, then a restriction of the population size to a single pair of individuals ('bottlenecking') should drastically affect selection limits since many of these alleles would not be represented in the sample of four gametes from which such a selection line was derived (Robertson, 1966a). Short-term results from bottleneck experiments (Da Silva, 1961; Frankham, 1980b) suggest that the variation in the initial population is mainly caused by alleles of an intermediate frequency. When selection has been continued over around 20 generations, the behaviour of the lines does no longer necessarily reflect the allelic frequencies in the base population but can be thoroughly changed by mutations arising de novo (Frankham 1980a; Hill, 1982). A further setback for multiple allelism is Robertson's (1980) speculation that the response upwards and downwards can be attributed to segregants within the same set of loci.

In addition to mutation, intragenic recombination or gene conversion (Watt, 1972; Strobeck and Morgan, 1977), and unequal crossing-over (Frankham, Briscoe and Nurthen, 1978) will generate new variation in a population, whereas selection and sampling will reduce it. Population genetics theory predicts (Ewens, 1972) that

if the mutation rate is u and the effective population size N_e, then given that $4N_eu = 1$, for sample size of 10, 50, 100, 200, the expected sample mean of the number of neutral alleles is 3.60, 5.19, 5.88, 6.57, respectively. Clearly, although our value for the $4N_eu$ cannot be regarded as too unrealistic if laboratory populations are considered, we need large experiments to study the properties caused by multiple alleles. Unless experiments are run at a very large scale, alleles are lost through sampling within few generations. Since our interest lies in the behaviour of the first few generations after sampling from the large base population, the assumption about the existence of a large number of variants is not too optimistic.

We shall investigate under what kind of circumstances the presence of multiple alleles may lead into non-linearity of offspring-parent regression, and also, how the degree of non-linearity is related to the number of alleles.

Let us assume that there are m alleles segregating at a locus affecting a quantitative trait. The allelic effects are assumed to form a uniform distribution between a and ma, so that an allele i ($i = 1, 2, \ldots, m$) has an effect $a_i = ia$ and frequency $q_i = 1/m$. Dominance is thought to act according to the ranked order of allelic effects, so that the value for a genotype comprising alleles i and j is

$$a_i + a_j + d|a_i - a_j| = a(i+j) + da|i-j|.$$

Hence we find for the overall mean

$$M = \Big(\sum_{i,j} (i+j) + 2d \sum_{i<j} (j-i) \Big) a/m^2 = (m+1)a + (m^2-1)da/3m.$$

At Hardy - Weinberg equilibrium the average effect of an allele i (α_i) can be calculated in a least-squares sense (Kempthorne, 1954), as

$$\alpha_i = \sum_{j=1}^{m} q_j a(i+j+d|i-j|) - M$$

$$= \{ \sum_{j=1}^{m} (i+j) + d \sum_{j=1}^{m} (i-j) + \sum_{j=i+1}^{m} (j-i) \} a/m - M$$

$$= (i+(m+1)/2)a + (2i^2-2i-2mi+m+m^2)da/2m - M. \qquad (24)$$

The dominance deviation due to alleles i and j is

$$\delta_{ij} = (i+j)a + da|j-i| - \alpha_i - \alpha_j - M.$$

The expressions, in terms of bivariate moments of additive and dominance deviations, for various expected values required in computing a quadratic regression, namely (16) and (18), were derived for a locus with an arbitrary number of alleles, and can therefore be used here. The moments are, in terms of average allelic effects and dominance deviations, for m alleles, as follows,

$$E(A^2) = 2 \sum_i q_i \alpha_i^2$$

$$E(D^2) = \sum_{i,j} q_i q_j \delta_{ij}^2$$

$$E(A^3) = 2 \sum_i q_i \alpha_i^3$$

$$E(D^3) = \sum_{i,j} q_i q_j \delta_{ij}^3$$

$$E(A^2 D) = \sum_{i,j} q_i q_j (\alpha_i + \alpha_j)^2 \delta_{ij}$$

$$E(AD^2) = \sum_{i,j} q_i q_j (\alpha_i + \alpha_j) \delta_{ij}^2$$

No attempt has been made to produce explicit formulae in terms of as, ds, and qs for these moments, instead a computer has been left to do the calculations.

The results for single parent and mid-parent regression for up to 10 alleles per locus are shown in Table 3. The degree of non-linearity caused by multiple alleles can be considered negligible within the model we have chosen, corresponding well with the case of two alleles at intermediate frequencies. When the number of alleles is increased the amount of curvature decreases, although the rate is not proportional to 1/m. The only conspicuous difference from a biallelic case is found when dominance is complete, namely the regression is no longer linear when there are more than two alleles segregating at a locus. Consider for simplicity a three-allele model, the genotypes B_1B_1, B_1B_2, and B_1B_3 having values 2a, B_2B_2 and B_2B_3 4a, and B_3B_3 6a. A straightforward calculation using (24) gives the regression of offspring on one parent, as

G-M	-10a/9	8a/9	27a/9
E(O|G)-M	-4a/9	4a/9	8a/9

Table 3

The relationship between non-linearity ($10*S_b$) of offspring-parent regression and number of alleles for various degrees of dominance at two equal loci.

(a) allelic effects equally spaced

		single parent			mid-parent			
d	m = 2	3	6	10	2	3	6	10
-1.0	0.00	-0.34	-0.41	-0.42	1.67	0.34	0.15	0.13
-0.9	-4.05	-0.46	-0.47	-0.46	1.07	0.22	0.10	0.08
-0.5	-1.25	-0.47	-0.35	-0.33	0.05	0.01	0.00	0.00
0.0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1.0	0.00	0.34	0.41	0.42	-1.67	-0.34	-0.15	-0.13

(b) allelic effects unequally spaced

a =2ma

		single parent			mid-parent			
d	m = 2	3	6	10	2	3	6	10
-1.0	0.00	-0.89	-2.88	-3.14	1.67	0.69	-0.51	-1.54
-0.9	-4.05	-1.80	-2.86	-2.56	1.07	0.25	-0.57	-1.17
-0.5	-1.25	-1.13	-0.86	-0.57	0.05	-0.00	-0.13	-0.14
0.0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.5	1.25	0.64	0.18	0.07	-0.05	-0.09	-0.06	-0.03
0.9	4.05	0.29	0.06	0.04	-1.07	-0.69	-0.30	-0.14
1.0	0.00	0.02	0.03	0.03	-1.67	-0.93	-0.38	-0.18

a =10ma

-1.0	0.00	-0.73	-2.81	-6.29	1.67	2.20	1.39	-0.11
-0.9	-4.05	-6.09	-10.87	-13.44	1.07	0.83	-0.30	-4.90
-0.5	-1.25	-1.55	-1.49	-1.18	0.05	-0.01	-0.29	-0.43
0.0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.5	1.25	-0.81	0.38	0.21	-0.05	-0.11	-0.09	-0.06
0.9	4.05	1.74	0.36	0.10	-1.07	-0.75	-0.37	-0.22
1.0	0.00	0.00	0.00	0.00	-1.67	-1.00	-0.45	-0.26

This regression is clearly non-linear, since the ratio of differences between successive values amongst parents is 6 : 5, but 2 : 1 amongst the expected progeny values.

When we assume that allelic effects follow a uniform distribution, we then essentially assume that differences between allelic contributions are relatively small. One way to study the consequences of increased heterogeneity amongst alleles is, for example, to give a value for the allele with the largest effect far apart from the rest. Two such cases have been considered, one when $a_m = 2ma$ and $a_m = 10ma$. Thus, for example, in the first case the series of four alleles have effects a, 2a, 3a, and 8a. The allele frequencies are assumed to be equal.

Table 3 shows that the larger the heterogeneity amongst alleles, the more the non-linearity resembles the one in two-allele case, which can be regarded as intuitively obvious. Hence when the difference between the most favourable allele and the others is very large, curvilinearity is most substantial when there is almost complete recessivity and several alleles, that is to say amongst all, also the outstanding allele is very rare. Furthermore, the single parent regression tends towards linearity when the dominance is complete. Obviously, symmetrical results would follow had we an allele with an extremely low value.

Normal Distribution Theory

The use of normal distributions for quantitative characters goes

back to the time of biometricians. During the two last decades the theory of quantitative genetics has concentrated on areas where normal distributions are inadequate, such as selection limits and effects of linkage. Recently, however, there has been a revival of their use, in quantitative genetics itself and in its application into evolutionary theory (Bulmer, 1971; Lande, 1976).

Under random mating the contributions from different loci, with small but similar effects, will be statistically independent, whether or not loci are linked, so that by the central limit theorem their sum will become asymptotically normal as the number of loci increases. This is often referred to as an infinitesimal model. Normality follows also, if there is a finite number of loci each with an infinite number of alleles whose effects on the measurement are normally distributed. Although normal distributions are chosen often for their mathematical tractability, their use can be also justified from their biological arguments.

From the bivariate form of the central limit theorem it follows that under random mating and in the absence of epistasis the joint genotypic distribution of parent and offspring will be approximately normal when the number of loci is large. In the absence of linkage a more generalised version can be stated, namely that for any number of related individuals the joint distribution of their genotypic values (and, obviously, their breeding values and dominance deviations, as well) will be multivariate normal, when the number of loci is very large. Given these statistical premises it follows immediately that amongst related individuals any regression of one of them on the rest is linear and the residual error about the

regression line is normally distributed with constant variance. As we have seen, when there are only few loci contributing to the variation of a character, dominance will usually cause departures from linearity. However, when the number of loci is large, linearity holds irrespective of dominance, although the slope of the regression line depends on the degree of dominance. If we write h_A^2 for the ratio of the additive variance to the total genotypic variance, we have under these assumptions for the genotypic regression of offspring on parent the expression

$$E(O|G) = M + h_A^2(G-M)/2,$$

where M stands for the overall mean, and from this, under random mating, for the mid-parental regression the expression

$$E(O|\bar{G}) = M + h_A^2(\bar{G}-M).$$

The conditional variances are found to be

$$V(O|G) = (1-h_A^4/4)\ V_G,$$

$$V(O|\bar{G}) = (1-h_A^4/2)\ V_G,$$

respectively.

This review of the normal distribution theory follows by and large Bulmer's (1976; 1980) treatment on the subject.

Few Major Loci beside an Infinite Number of Minor Loci

It may not be an exaggeration to suggest that every locus contributes in one way or another to any given quantitative character, although the effects of most loci are vanishingly small. It seems realistic to assume that the contributions of the loci are vastly unequal, there being a small number of major loci and a larger number of minor loci (Robertson, 1967). For example, in wheat Wehrhahn and Allard (1965), using a technique of backcrossing followed by self-fertilization, performed an analysis of the variation in heading time and showed that 94% of the additive variance in the trait accounted by only four loci. Similar kind of results have been obtained for Drosophila bristle characters (e.g. Thoday, Gibson and Spickett, 1964; Frankham and Nurthen 1981). Genes with considerable effects on a metric character have been found also in some domestic animals, e.g. a gene for high fecundity in Booroola strain of Merino sheep (Piper and Bindon, 1982).

Various tests have been suggested to detect major locus effects either by studying the shape of various distributions (Merat, 1968; Hammond and James, 1970) or the variance of the offspring from given parents (e.g. Matthysse, Lange and Wagener, 1979). As well, a number of quantitative trait loci are claimed to have been mapped (e.g. Thoday, 1979; Shrimpton, 1982). Frankham et al.(1981) proposed a test for rare alleles of large increasing or decreasing effect, based on the changes in genetic variation under artificial selection, into situations where an estimate of environmental variance is easily available. The discriminatory power of these methods seems to be very low, and possibly thousands of observations

are required to detect such major loci (e.g. McMillan and Robertson, 1974).

Let us assume that the genetic variation in a trait is caused by n major loci and an infinite number of loci with small effects, the gene action being additive between loci. The parameters describing the residual variation, i.e. the variation accounted by other than the major loci, are denoted by a subscript R and the ones attributed to the major loci, by n. From the normal distribution theory we know that the quadratic coefficient equals zero in both single and mid-parent regressions when the variation is due to a large number of loci. Since then $\mu_{3G} = 0$, we find from (20) that $E(og^2) = E(og_m g_f) = 0$. We allow for dominance amongst the infinite number of loci, and denote the additive proportion of the residual genotypic variance by h_R^2. Thus, in general, h_R^2 does not equal one.

It was shown earlier by (21) that when the variation is solely due to n biallelic loci with complete dominance then the single parent regression is linear given equal allelic frequencies over loci. In general, this ceases to be true if these loci act together with an infinite number of loci. Since for the normal distribution, cumulants of third or higher order are all zero, we find that the quadratic coefficient of single parent regression can be written, as

$$b_2 = \frac{\{V_{Gn}+V_{GR}\}E(og^2)_n - \mu_{3Gn}\{h_n^2 V_{Gn}+h_R^2 V_{GR}\}/2}{\{\mu_{4Gn} - 3V_{Gn}^2 + 2(V_{Gn}+V_{GR})^2\}\{V_{Gn}+V_{GR}\} - \mu_{3Gn}^2} \qquad (25)$$

In the case of complete dominance we have for the major loci, from

(21), as

$$V_{Gn} E(og^2)_n - \mu_{3Gn} h^2_n V_{Gn}/2 = 0.$$

The numerator of (25) is therefore

$$V_{GR} \mu_{3Gn} (h^2_n - h^2_R)/2. \qquad (26)$$

This is seen to equal zero, except in trivial case of $V_{GR} = 0$, if $\mu_{3Gn} = 0$ or $h^2_n - h^2_R$. When there is complete dominance the distribution is symmetrical (or $\mu_{3Gn} = 0$), i.e. for $d = -1$, $q = 1/\sqrt{2}$ and for $d = 1$, $q = 1 - 1/\sqrt{2}$. The other condition for linearity is when there is the same proportion of additive variation in the two different types of loci. Let us consider a case when $d = -1$. If now $q < 1/\sqrt{2}$, μ_{3Gn} is positive and the numerator is negative, its value being proportional to q, unless most of the residual variation is of a non-additive kind.

In an analogous way to (25) we find for the mid-parent regression

$$b_{\bar{2}} = \frac{4\{V_{Gn}+V_{GR}\}\{E(og^2)_n+E(og_m g_f)_n\} - 2\mu_{3Gn}\{h^2_n V_{Gn}+h^2_R V_{GR}\}}{\{\mu_{4Gn}-3V_{Gn}+4(V_{Gn}+V_{GR})^2\}\{V_{Gn}+V_{GR}\} - \mu^2_{3Gn}} \qquad (27)$$

The results for the model have been given in Table 4 for various levels of h^2_R and for three sizes of gene effect expressed relative to the standard deviation of the residual variation when $n = 1$. Only the lower range of q is represented, because the degree of non-linearity is symmetrical in d and q. When at the major loci

Table 4

Non-linearity ($10*S_b$) in offspring-parent regression whem there is one major locus acting in a background variation due to a large number of loci with small effects. Various gene frequencies, degrees of dominance, magnitudes of gene effects and proportions of additive variance in the residual variation have been considered.

$a/\sigma_R = 0.5$

			single parent			mid-parent	
q	d	$h^2 =$ 1.0	0.5	0.1	1.0	0.5	0.1
0.05	-1.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00
	-0.90	-0.00	-0.00	0.00	-0.00	-0.00	0.00
	-0.50	-0.00	0.00	0.00	-0.00	0.00	0.00
	0.00	0.00	0.00	0.03	0.00	0.00	0.01
	0.50	-0.00	0.00	0.15	-0.00	0.00	0.07
	0.90	-0.00	0.01	0.36	-0.00	0.00	0.16
	1.00	-0.00	0.01	0.43	-0.00	0.00	0.18
0.10	-1.00	-0.00	-0.00	0.00	-0.00	-0.00	0.00
	-0.90	-0.00	-0.00	0.00	-0.00	-0.00	0.00
	-0.50	-0.00	0.00	0.01	-0.00	0.00	0.01
	0.00	-0.00	0.00	0.05	0.00	0.00	0.03
	0.50	0.00	0.01	0.17	-0.00	0.00	0.07
	0.90	-0.00	0.01	0.31	-0.00	0.00	0.11
	1.00	-0.00	0.01	0.35	-0.00	0.00	0.12
0.50	-1.00	-0.00	0.00	0.17	0.00	0.01	0.24
	-0.90	-0.00	0.00	0.14	0.00	0.01	0.20
	-0.50	-0.00	0.00	0.04	0.00	0.00	0.06
	0.00	0.00	0.00	0.00	0.00	0.00	0.00
$a/\sigma_R = 1.0$							
0.05	-1.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00
	-0.90	-0.00	-0.00	0.00	-0.00	-0.00	0.00
	-0.50	-0.00	0.00	0.07	-0.00	0.00	0.04
	0.00	0.00	0.02	0.58	0.00	0.01	0.29
	0.50	0.00	0.09	1.51	-0.00	0.04	0.70
	0.90	-0.00	0.16	2.15	-0.00	0.06	0.95
	1.00	-0.00	0.18	2.27	-0.00	0.07	0.99
0.10	-1.00	-0.02	-0.01	0.01	-0.01	-0.00	0.01
	-0.90	-0.01	-0.01	0.04	-0.01	-0.00	0.03
	-0.50	-0.00	0.00	0.25	-0.00	0.00	0.17
	0.00	0.00	0.03	0.64	0.00	0.02	0.32
	0.50	0.00	0.09	0.98	-0.00	0.02	0.37
	0.90	-0.00	0.12	1.09	-0.01	0.02	0.31
	1.00	-0.00	0.12	1.10	-0.01	0.02	0.29
0.50	-1.00	-0.03	0.02	0.53	0.01	0.16	1.02
	-0.90	-0.03	0.01	0.41	0.00	0.12	0.82
	-0.50	-0.02	0.00	0.11	0.00	0.03	0.25
	0.00	0.00	0.00	-0.00	0.00	0.00	0.00

Table 4 (cont.)

$a/\sigma_R = 5.0$

q	d	h^2 =	single parent 1.0	0.5	0.1	mid-parent 1.0	0.5	0.1
0.05	-1.00		-1.69	-1.29	-0.00	-1.47	-1.11	0.00
	-0.90		-1.28	-0.74	0.16	-1.07	-0.58	0.27
	-0.50		-0.19	0.00	0.83	-0.09	0.03	0.94
	0.00		0.00	0.33	1.43	-0.00	0.19	0.83
	0.50		0.00	0.28	0.99	-0.01	0.04	0.24
	0.90		-0.00	0.17	0.62	-0.05	0.00	0.04
	1.00		-0.00	0.15	0.55	-0.07	-0.00	0.02
0.10	-1.00		-4.23	-1.93	0.75	-3.39	-1.35	1.30
	-0.90		-3.17	-1.28	0.18	-2.36	-0.74	0.56
	-0.50		-0.51	-0.10	0.03	-0.21	-0.00	0.18
	0.00		-0.00	0.11	0.42	0.00	0.05	0.20
	0.50		0.01	0.18	0.46	-0.04	-0.00	0.00
	0.90		0.00	0.07	0.24	-0.15	-0.09	-0.05
	1.00		-0.00	0.05	0.19	-0.19	-0.13	-0.08
0.50	-1.00		-0.03	0.01	0.09	1.08	1.36	1.63
	-0.90		-0.27	-0.09	-0.01	0.69	0.89	1.09
	-0.50		-0.65	-0.51	-0.41	0.03	0.06	0.09
	0.00		0.00	0.00	-0.00	0.00	0.00	-0.00

d is less than -0.5 and the gene frequency low, $E(og^2)$ and $E(og_m g_f)$ are small compared with μ_{3G} (Figure 3). The recessive alleles at low frequency contribute very little additive variance, we may conclude that the numerator of quadratic coefficient, in both regressions, is roughly equal to $-h_R^2 \mu_{3Gn} V_{GR}$. In this case the largest deviations from linearity can be expected when d tends to -1 and q to 0, the curve being curved downwards. Since (26) holds also when d = 0, we notice that for gene frequencies apart from 0.5 the regression is linear only if $h_R^2 = 1$, and for $h_R^2 < 1$ curved upwards at low frequencies and downwards at high frequencies. Similarly, we find for the mid-parent regression that when d at the major loci equals either 0 or ± 1, the numerator of (27) differs under these conditions from (26) only by a factor two , and hence the results obtained for single parent regression above hold here, as well.

If h_R^2 is small compared with h_n^2, the numerator in (25) is roughly equal to $h_n^2 V_{Gn} \mu_{3Gn}/2$, when $q < 0.5$ and $d > 0$, since the genotypic regression is then almost linear (cf. Figure 4).

The degree of non-linearity depends also on the proportion of variation attributed to the major loci, that is upon a/σ_R, so that when a/σ_R is very small we have linearity and when very large the non-linearity is naturally best appreciated from Figure 4.

Summary

i. The degree of non-linearity in the regression between the

genotypic values of offspring and parent has been studied using quadratic regression when a character is determined by additive loci with dominance.

ii. The third moment of genotypic distribution and the covariances of offspring mean with the squared parental values and with the product of the two are derived in terms of bivariate moments of additive and dominance deviation for a locus with an arbitrary number of alleles. The third moment and the covariances, and consequently the amount of non-linearity in regression, are in their absolute magnitude symmetrical with respect to gene frequency and dominance.

iii. When the variation is caused by a small number of loci each with the same effect, gene frequency and dominance, the amount and type of curvature (+ for convex, - for concave, and 0 for linear regression) with respect to frequency and dominance relationship of the allele with an increasing effect is for the single parent regression as follows:

	low	intermediate	high
completely recessive	0	0	0
partially recessive	--	-	∿ 0
additive	0	0	0
partially dominant	∿ 0	+	++
completely domiannt	0	0	0

and for the mid-parent regression as follows:

	low	intermediate	high
completely recessive	+	+	∿ 0
partially recessive	--	+	∿ 0
additive	0	0	0
partially dominant	∿ 0	-	++
completely dominant	∿ 0	-	-

iv. In relation to the number of alleles largest deviations from linearity occur when there are only two alleles segregating per locus, non-linearity decreasing when the number of alleles increases. When the number of loci conributing to the variation is increased non-linearity decreases, the rate being roughly proportional to 1 / number of loci. When the number is very large, genotypic distribution is normal and the regression linear.

v. When the variation is caused by loci with similar effect but different gene frequency and dominance the non-linearity is roughly the same as with equal loci considering directional dominance and averaged gene frequency over loci. The regression is linear and genotypic distribution symmetrical when there are pairs of loci where the dominance relationship and gene frequency between alleles is interchanged.

vi. When there is a locus with a large effect acting in a genetic background due to an infinite number of loci, largest departures from linearity are caused by a rare recessive allele, if most of the residual variation is additive, or by a rare dominant allele if non-additivity makes substantial contribution to the residual variation.

4. GENOTYPIC REGRESSION UNDER MULTIPLICATIVE MODEL

Introduction

Besides dominance, epistasis is another kind of non-additivity, in the presence of which the predictions based on an overall heritability are likely to break down. There is a particular type of epistasis, multiplicative interaction between loci, which can be removed by transformation. It may be called a metrical or scale effect, since it depends on the scale used for measuring. The 'right' scale in terms of our prediction equation would be the one on which the effect of a given gene substitution is constant, irrespective of the genetic background on which it is made, i.e. the logarithmic scale. An immediate difficulty arises, transformation is done on the phenotypic distribution assuming that genetic and non-genetic effects act on the same scale. There is some evidence from, e.g. the work of Powers (1950) that in some situations the genetic variation requires one scale, while the environmental variation requires another. If the alleles within a locus do not interact in a multiplicative way - and there is no such principle why they should - the log transformation alters the type of allelic interaction towards dominance of the allele with a larger effect (e.g. Lerner, 1958). However, beside the practical problem of knowing when and how to transform a given body of data, it is of interest to learn how serious the consequences of not transforming may be when this type of gene action is present.

In the additive model we used the properties of cumulants when we

calculated the moments of the genotypic variation accounted for several loci. For the multiplicative model these central moments are calculated starting from the moments about zero of the single locus contributions. Since the loci are assumed to be independent of each other (i.e. no linkage disequilibrium), the moment of any order about zero for n loci is a product of the corresponding moments of single locus contributions over n loci, or with equal loci simply the nth power of the moment for one locus. We can write various moments about zero for the contributions from a locus i as

$$
\begin{aligned}
&E(G_i) = m, \\
&E(G_i^2) = m_2 + m^2, \\
&E(G_i^3) = m_3 + 3mm_2 + m^3, \\
&E(O_i G_i) = m_2 + E(A^2)/2, \\
&E(O_i G_i) = m_3 + mm_2 + mE(A^2) + E(A^3)/2 + E(A^2 D) + E(AD^2)/2, \\
&E(O_i G_{mi} G_{fi}) = m_3 + mE(A^2) + E(A^2 D)/2,
\end{aligned}
\tag{26}
$$

where m is the mean, m_r the rth moment about m of the contributions from this locus, and the moments of A and D are defined as in the additive model ignoring interaction effects from other loci. Under the multiplicative model the genotypic value for n loci is $\prod_{i=1}^{n} G_i$, and $E(\prod_i G_i) = \prod_i E(G_i)$, $E(\prod_i O_i G_i^r) = \prod_i E(O_i G_i^r)$, for $r = 1, 2, 3, \ldots$, and $E(\prod_i O_i G_{mi} G_{fi}) = \prod_i E(O_i G_{mi} G_{fi})$.

The importance of the scale transformation is very much dependent on the degree of variation in the character we are working with. Scale transformation may be of little value if we are working within a population with a small coefficient of variation. We want to choose our model - or more precisely, the difference between two homozygotes at a locus - to be consistent with the possible range in the magnitude of the genotypic variance. The genotypic coefficient of variation (denoted by C) usually will not exceed 40% and the environmental variance is, in general, more than one half of the phenotypic one. Thus, a reasonable upper limit for the genotypic coefficient of variation is $40/\sqrt{2} \sim 30\%$. In our calculations we consider effect of multiplicative interaction between loci at two levels of variation, namely when the coefficient of genotypic variation is low ($\sim 5\%$) and high ($\sim 30\%$).

Models with Equal Loci

Let us assume that there are n equal loci, with two alleles at each locus, contributing to the variation. The genetic model is written as follows,

genotype	bb	Bb	BB
value	1	1+(1+d)a	1+2a
frequency	$(1-q)^2$	$2q(1-q)$	q^2

mean = $1 + 2aq(1+d(1-q))$

Since the difference between the two homozygotes at an individual locus is 2a and the dominance relationship is defined as in the additive model, we can use the formulae derived for that model.

In the multiplicative model the coefficient of the genotypic variation determined by n equal loci is, in terms of single locus moments the following expression

$$C = \sqrt{(E(G^2)^n - E(G)^{2n})}/E(G)^n,$$

hence

$$(C^2 + 1)^{1/n} = E(G^2)/E(G)^2$$

$$= (3a^2 + 2a + 3)/(2a^2 + 4a + 2),$$

when $d = 0$ and $q = 0.5$, from which a can be found in terms of n and C. From the two solutions of a, one larger and one smaller than zero, only the former, $a = (2k-1+2\sqrt{(2k-2)})/(3-2k)$, where $k = (C^2+1)^{1/n}$, rendering positive interaction is considered. There is good evidence from Drosophila which is almost the only organism where a detailed analysis of the interactions between genetic factors can be satisfactorily done, that the genetic factors controlling simple quantitative characters, such as the number of sternopleural bristles, interact with each other positively (Robertson, 1970). With less variation or more segregating loci a would be necessarily smaller. When $n = 2$, we find, for

$$C = 0.05,\ 0.10,\ 0.20,\ 0.30,$$

$1+2a = 1.11, 1.22, 1.50, 1.84,$

respectively, and when $n = 20$, we have for

$$C = 0.05, 0.10, 0.20, 0.30,$$

$$1+2a = 1.03, 1.07, 1.13, 1.20.$$

The use of coefficient of variation as a basis to choose the value for a in the multiplicative model was first suggested by Horner, Comstock, and Robinson (1955). From (26) we obtain the variance for n loci in terms of single locus moments as

$$V_G = (m+m_2)^n - m^{2n}$$

$$= m^{2n}(nm_2/m^2 + n(n-1)m_2^2/2m^4 + \ldots + m_2^n/m^{2n}).$$

We notice that allelic interactions between loci (terms apart from nm_2/m^2) contribute relatively little into the total genotypic variance, the contribution being proportional to $\sqrt{m_2}/m$, the coefficient of variation of the effects from individual loci, and hence to the total genotypic coefficient of variation.

The consequences of epistatic gene action was made precise by the pioneering work of Cockerham (1954), and of Anderson and Kempthorne (1954), who showed how the epistatic quantitative variation could be partioned and how the components contribute to the similarities between relatives. Because of small coefficients of these components in most of the observable variances, there are very few

situations in which epistasis can be regarded as important. In the presence of epistatic interactions the offspring-parent covariance, whether referring to one parent or mid-parent, equals $V_A/2 + V_{AA}/4 + V_{AAA}/8 + \ldots$, where V_A is the additive variance, and V_{AA} and V_{AAA} variances due to two- and three-locus interactions between breeding values. Interactions involving more loci contribute progressively smaller proportions as the order of interactions increases. The effect of the interaction components is that the offspring-parent covariance is more than half of the additive variance. When the interaction between loci is of a multiplicative type we have for n equal loci, from (26),

$$\mathrm{Cov}(O,G) = m^{2n}(nE(A^2)/2m^2 + n(n-1)E(A^2)/8m^4 + \ldots + E(A^2)/2^n m^{2n})$$

Terms, apart from the one referring to single locus contributions additive over loci, are seen to be negligible. For example, when there are 20 loci, with $d = 0$ and $q = 0.5$, twice the covariance is only 2% larger than the additive variance ($V_A = nm^{2n-2}E(A^2)$), when $C = 0.30$, the difference being smaller with fewer loci and smaller coefficient of variation, respectively. Similar conclusions can be made about $\mathrm{Cov}(O,G^2)$ and $\mathrm{Cov}(O,G_m G_f)$.

Because of the positive interaction we expect the genotypic distribution, in general, to tail off to the right. Hence symmetry of third order moments, with respect to d and q, found in the additive case cannot be met in the multiplicative. When the plus allele is recessive and at a low frequency the interaction will reinforce the positive skewness of the genotypic distribution. On the other hand when the dominance relationship and frequencies

Table 5

Symmetry of the skewness and kurtosis of genotypic distribution with respect to gene frequency and degree of dominance when the variation is due to different number of loci with either additive or multiplicative gene action between loci.

a) skewness of genotypic distribution

model			additive		multiplicative			
					low C		high C	
q	d	n =	2	20	2	20	2	20
0.1	1.0		6.96	2.20	6.99	2.24	7.27	2.50
	-0.5		2.77	0.88	2.80	0.93	3.08	1.26
	0.0		1.33	0.42	1.38	0.51	1.76	1.01
	1.0		1.12	0.35	1.21	0.52	1.90	1.50
0.5	-1.0		0.82	0.26	0.91	0.44	1.58	1.50
	0.0		0.00	0.00	0.07	0.14	0.43	0.87
	1.0		-0.82	-0.26	-0.91	-0.44	-1.58	-1.50
0.9	-1.0		-1.12	-0.35	-1.04	-0.20	-0.80	0.49
	0.0		-1.33	-0.42	-1.29	-0.34	-1.15	0.03
	0.5		-2.77	-0.88	-2.74	-2.04	-2.66	-0.60
	1.0		-6.96	-2.20	-6.95	-2.16	-6.90	-2.01

b) kurtosis of genotypic distribution

model			additive		multiplicative			
					low C		high C	
q	d	n =	2	20	2	20	2	20
0.1	-1.0		47.51	4.75	48.18	5.12	57.29	7.68
	-0.5		10.91	1.09	11.26	1.29	14.95	2.75
	0.0		1.28	0.13	1.54	0.29	3.80	1.76
	1.0		0.25	0.02	0.66	0.30	4.11	4.20
0.5	-1.0		-0.33	-0.03	-0.01	0.21	2.41	4.19
	0.0		-0.50	-0.05	-0.50	-0.01	-0.35	1.29
	1.0		-0.33	-0.03	-0.61	-0.16	-1.47	0.50
0.9	-1.0		0.25	0.02	-0.09	-0.15	-1.07	-0.01
	0.0		1.28	0.13	1.06	-0.00	0.35	-0.29
	0.5		10.91	1.09	10.62	0.91	9.80	0.28
	1.0		47.52	4.75	46.99	4.41	45.75	3.17

between alleles are interchanged, the scale effect shifts the distribution towards symmetry. Obviously, the effect of multiplicative interaction is more noticeable when C is higher and/or the number of loci is large. In Table 5 skewness have been calculated for various degrees of dominance and levels of interactions and for a small and large number of loci.

While skewness is used to illustrate the degree of symmetry of a distribution, kurtosis, γ_2 , is used as a measure of relative peakedness (or flatness) of a distribution. It is a standardised fourth moment, usually defined as $\gamma_2 = \mu_4 / \mu_2^2 - 3$, taking values between -2 and infinity (the normal distribution has a kurtosis of zero). Distributions with a kurtosis greater than 0 are described as leptokurtic (with a high, narrow peak) and distributions with a kurtosis less than 0 as platykurtic (with a low, broad peak). The multiplicative interaction shifts distributions, which are positively skewed due to single locus conributions, towards leptokurtism and distributions which are negatively skewed, towards platykurtism, as shown in Table 5. The effects are again more conspicuous, when there is a large number of loci segregating and/or the coefficient of variation is high.

The amount and type of non-linearity expressed in a similar way as in the additive model has been computed for the genotypic regression of offspring on parent at two levels of interaction when the number of loci affecting a trait is four and 1+2a equals 1.073 or 1.527 corresponding to a low (C$\sim$0.05) and high (C$\sim$0.30) degreee of variation, respectively The results for various gene frequencies are plotted against degree of dominance in Figure 6a. Since

Figure 6

The relationship between non-linearity in offspring-parent regression and degree of dominance for five different gene frequencies when the variation is due to four equal multiplicative loci with (a) 1+2a=1.073 (C∿0.05), and (b) 1+2a=1.527 (C∿0.30). Solid line refers to single parent and broken line to mid-parent regression.

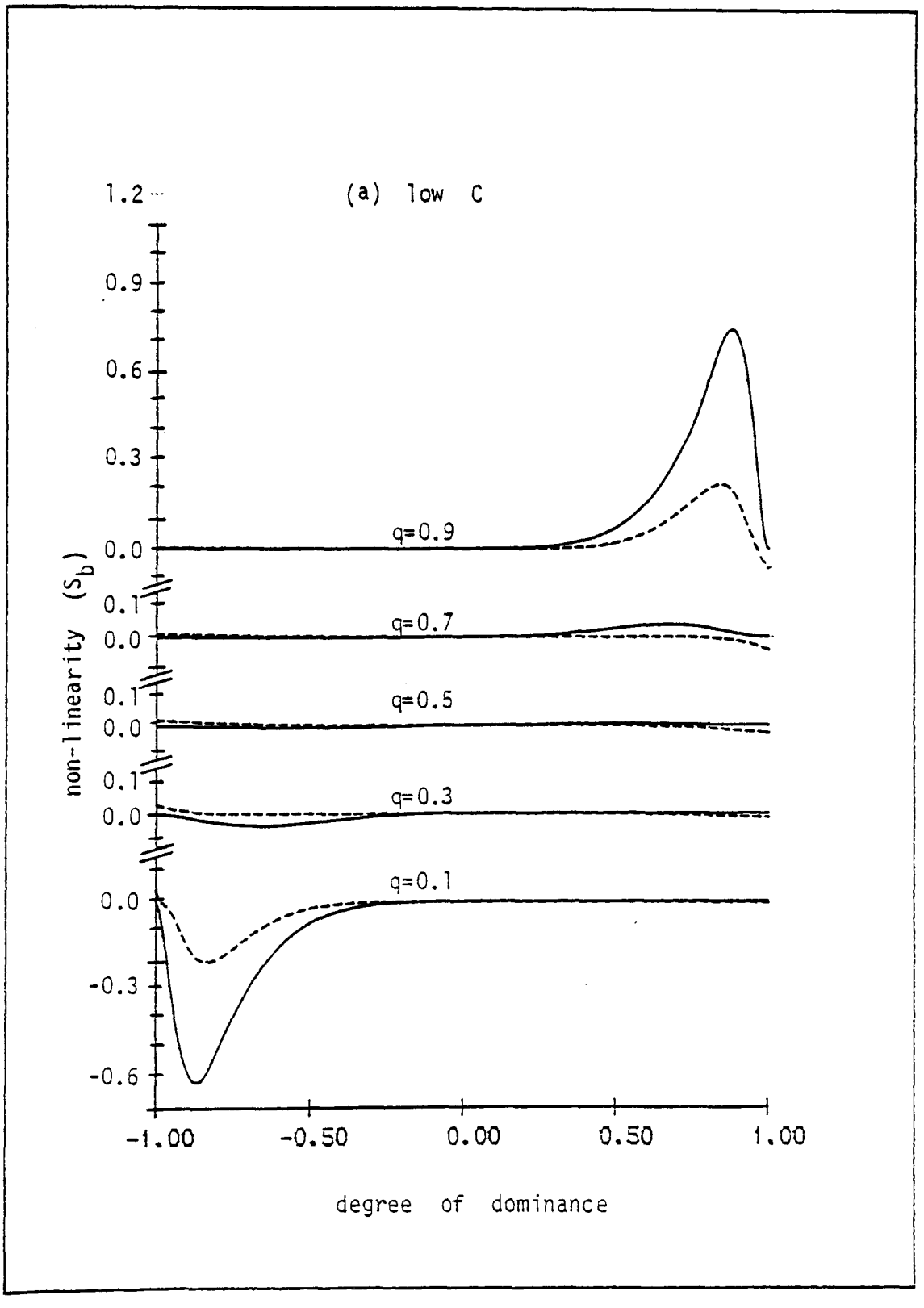
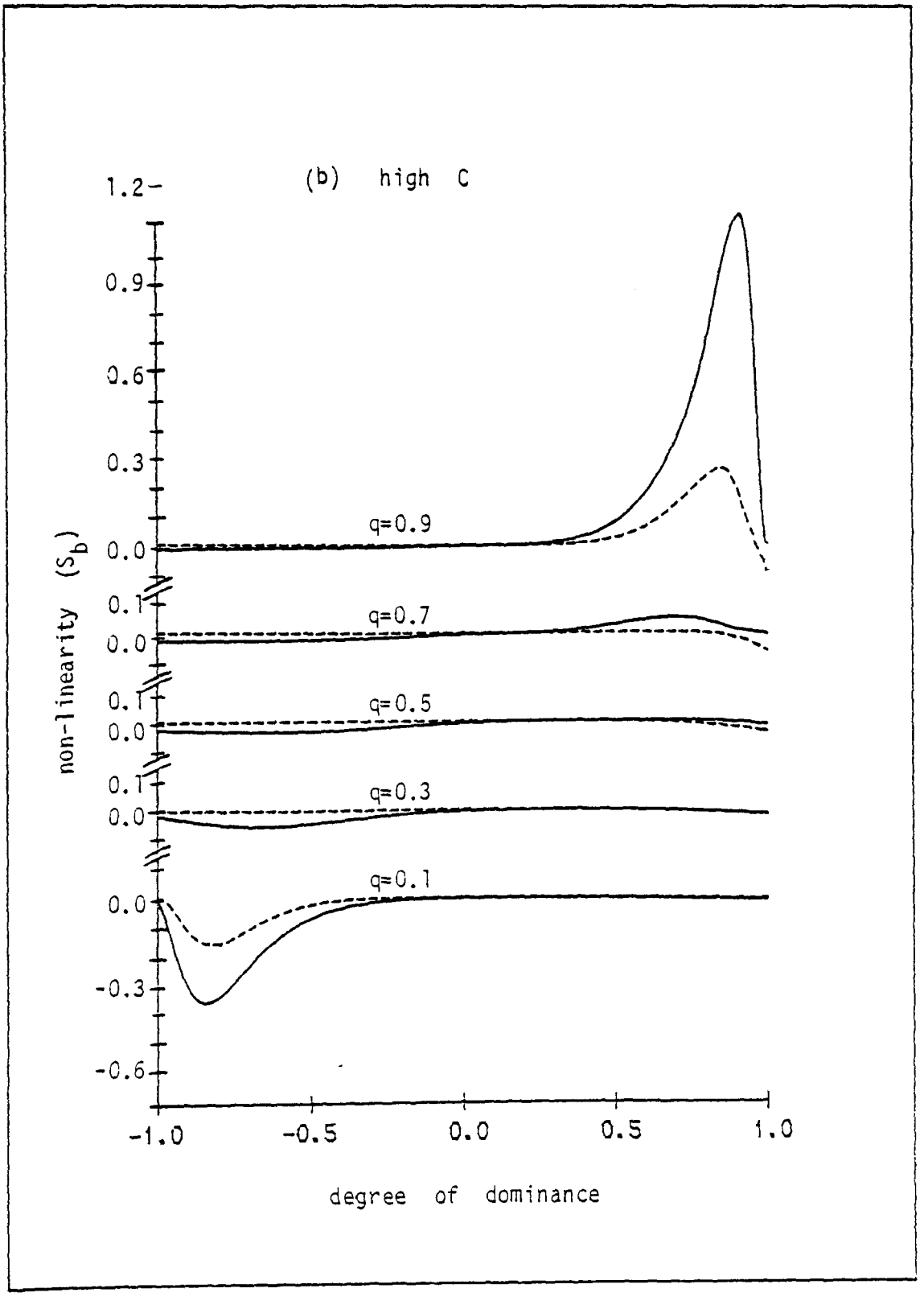


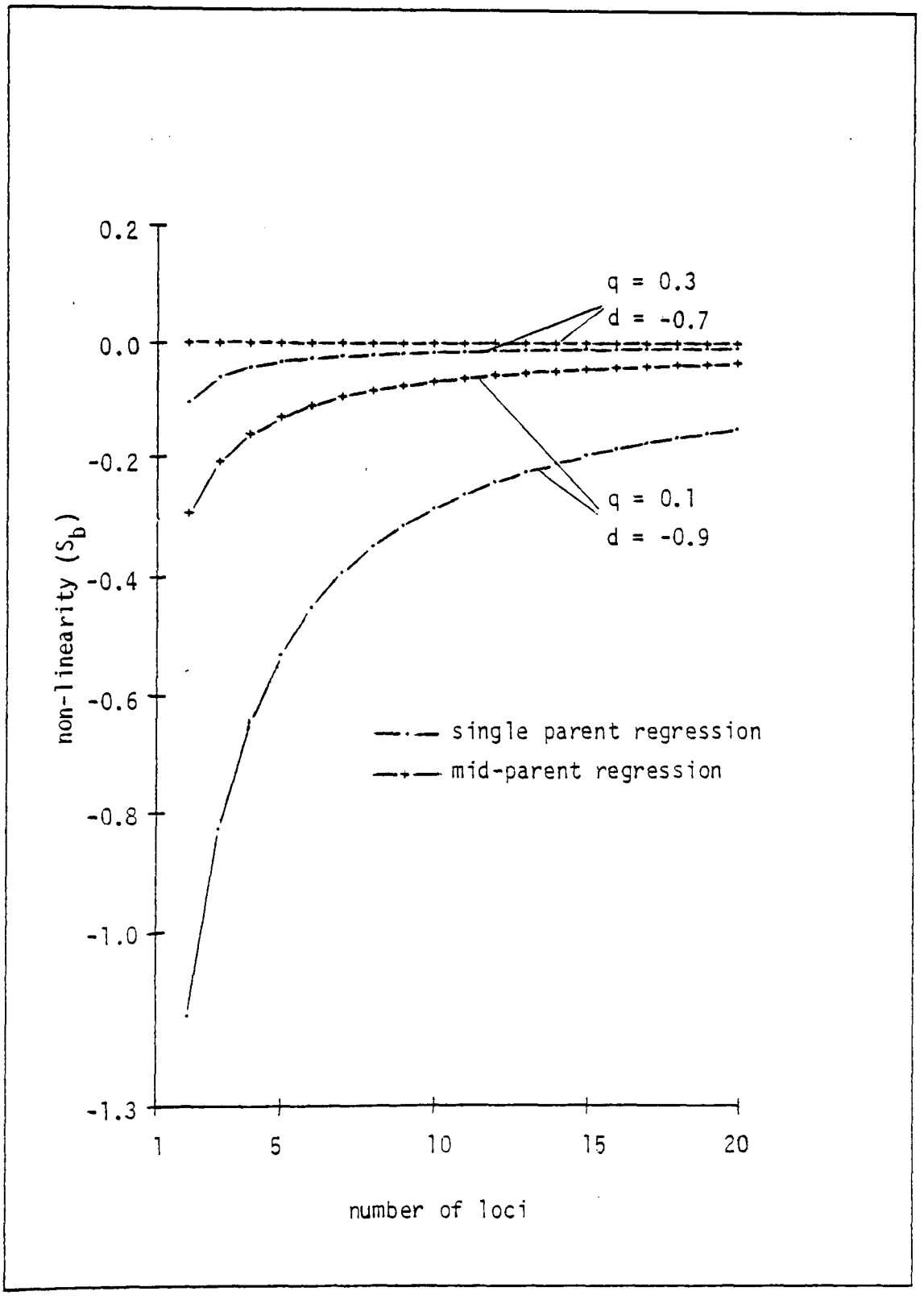
Figure 6 (cont.)



multiplicative interaction makes relatively small contributions into various moments, especially into the covariances, it is not surprising to find that the inclusion of that kind of interlocus effects does not substantially change the pattern of curvilinearity caused by dominance. As in the additive case, the regression on one parent is seen to be curved downwards when the plus allele is recessive, and upwards when it is dominant. Furthermore, the most noticeable deviations from linearity arise from rare, almost completely recessive alleles, so that given $d < 0$ the concavity is at its maximum when the allele frequency is around $1+d$, and for $d > 0$ when q is roughly equal to d.

When, with higher degree of variation, interaction effects are more substantial, the differences between additive and multiplicative models become more conspicuous (Figure 6b). In the cases where single parent regression is linear in the additive case, i.e. when $d = 0$ or ± 1, it is now slightly concave. Due to the asymmetry in the moments of genotypic distribution (cf. Table 5), we have a clear difference between the two gene action models, namely under multiplicative gene action there are considerable departures from the symmetry in the amount of non-linearity with respect to q and d. Dominance of an allele with a higher value and multiplicative interaction between loci are seen to augment one another in their contributions to curvilinearity, whilst the case is opposite when the allele is recessive. However, an unusually high coefficient of variation is obviously required to construct a model to overrule dominance as the main source of genotypic non-linearity. Figure 7 shows that when the number of loci affecting the variation is increased the non-linearity decreases the rate being roughly

Figure 7
The relationship between genotypic non-linearity of offspring-parent regression and number of loci for a given gene frequency and degree of dominance under multiplicative gene action (high coefficient of variation).



proportional to 1/n.

Multiple Alleles

The extension of multiplicative model to deal with multiple alleles is fairly straightforward. Suppose that there are m alleles segregating at a locus affecting a quantitative character. As in the additive model we assume that the allelic effects form a uniform distribution, in this case between 1/2 and 1/2 + (m-1)a (where 1/2 is merely a location parameter). An allele i (i= 0, 1, ..., m) has an effect a_i = ia and a frequency q_i = 1/m. Dominance is defined as in the additive case, so that the contribution from a locus containing alleles i and j is 1 + a(i+j) + d|i-j|. Since the difference between values of the two extreme genotypes and the dominance relationship of alleles are similar to those in the corresponding additive model, we can proceed as in the two-allele case in producing the central moments from the single locus moments about zero.

Two levels of interactions have been considered, the low one where the contributions to a genotypic value from a locus cover the range between 1 and 1.1052 for n = 2, and the high one where this range is from 1 to 1.8439 for n = 2 and from 1 to 1.3030 for n = 10, the levels corresponding to values 0.05 and 0.30 of the coefficient of genotypic variation arising from the given number of biallelic loci. The results for single parent and mid-parent regression for up to 10 alleles are in Table 6. The inclusion of multiplicative interaction between loci does not make

Table 6

The relationship between non-linearity ($10*S_b$) of offspring-parent regression and number of alleles for various degrees of dominance when loci are interacting multiplicatively. [a] n = 2 and low C, [b] n = 2 and high C, and [c] n = 10 and high C.

	single parent				mid-parent			
d	m = 2	3	6	10	2	3	6	10
-1.0	-0.00	-0.12	-0.16	-0.16	0.50	0.13	0.06	0.05 [a]
	-0.20	-0.32	-0.32	-0.30	0.22	0.06	0.03	0.03 [b]
	-0.27	-0.24	-0.20	-0.18	-0.00	-0.00	-0.00	-0.00 [c]
-0.9	-0.05	-0.16	-0.18	-0.18	0.32	0.08	0.04	0.03
	-0.29	-0.35	-0.33	-0.31	0.13	0.03	0.02	0.02
	-0.28	-0.24	-0.19	-0.17	-0.00	-0.00	-0.00	-0.00
-0.5	-0.23	-0.17	-0.14	-0.14	0.02	0.00	0.00	0.00
	-0.46	-0.32	-0.25	-0.23	0.00	0.00	0.00	0.00
	-0.26	-0.19	-0.14	-0.13	-0.00	-0.00	-0.00	-0.00
0.0	-0.00	-0.00	-0.00	-0.00	0.00	0.00	0.00	0.00
	-0.03	-0.02	-0.01	-0.01	0.00	0.00	0.00	0.00
	-0.09	-0.06	-0.04	-0.04	-0.00	-0.00	-0.00	-0.00
0.5	0.19	0.14	0.12	0.11	-0.02	-0.01	-0.00	-0.00
	0.18	0.09	0.07	0.07	-0.02	-0.00	-0.00	-0.00
	-0.02	-0.01	-0.00	-0.00	-0.01	-0.00	-0.00	-0.00
0.9	0.01	0.11	0.14	0.15	-0.39	-0.10	-0.04	-0.04
	-0.00	0.07	0.09	0.10	-0.47	-0.11	-0.05	-0.04
	-0.14	-0.03	-0.01	-0.01	-0.13	-0.04	-0.02	-0.02
1.0	-0.00	-0.08	-0.12	-0.13	-0.61	-0.15	-0.07	-0.06
	-0.11	0.04	0.08	0.08	-0.76	-0.18	-0.08	-0.07
	-0.22	-0.04	-0.01	-0.01	-0.21	-0.06	-0.03	-0.02

the pattern of curvilinearity much different from the corresponding additive model. In general, with larger interaction effects, caused either by higher degree of variation or by more loci contributing to the variation, the non-linearity tends towards low concavity irrespective of the number of alleles. This corresponds to the case of two alleles with intermediate frequencies, where consequently the single locus contributions are distributed fairly symmetrically despite dominance.

Infinite Number of Loci

When multiplicative interaction is removed by logarithmic transformation, the genotypic value determined by n loci can be written as $\sum_{i=1}^{n} \ln G_i$, where G_i is the contribution from the ith locus. When there is linkage equilibrium in the population, the distribution of genotypic values on log scale, given small but similar effects over loci, tends to a normal distribution when the number of loci contributing to the variation becomes very large, hence the untransformed values follow a so-called lognormal distribution. If μ and σ^2 are the mean and variance of the underlying normal distribution and the corresponding lognormal distribution has a density function f(G), we have for the rth moment of G about zero,

$$E(G^r) = \int_0^\infty G^r f(G)\, dG,$$

which after writing $u = (\ln G - \mu)/\sigma$ becomes

$$E(G^r) = \frac{1}{\sqrt{2\pi}} \int \exp\{r(u\sigma+\mu)\} \exp\{-u^2/2\}\, du$$

$$= \exp\{r\mu + r^2\sigma^2/2\}.$$

Hence

$$M = \exp\{\mu + \sigma^2/2\},$$

$$V_G = M^2(\exp\{\sigma^2\} - 1). \tag{27}$$

It is convenient to use the coefficient of variation for the lognormal distribution, since the skewness and kurtosis can be written as

$$\gamma_1 = C(3 + C^2),$$

$$\gamma_2 = C^2(16 + 15C^2 + 6C^4 + C^6),$$

respectively, and from (27) we have

$$C^2 = \exp\{\sigma^2\} - 1.$$

Thus the shape of the distribution depends only of its coefficient of variation (or of the variance, but not of the mean, of the underlying scale). From the formulae for γ_1 and γ_2, we notice that the lognormal distribution is always positively skewed and leptokurtic, e.g. for

$$C = 0.05,\ 0.15,\ 0.30$$

we have

$$\gamma_1 = 0.15,\ 0.45,\ 0.93$$

$$\gamma_2 = 0.04,\ 0.37,\ 1.57,$$

respectively. For the underlying distribution we find, in terms of parameters describing lognormal distribution, as

$$\sigma^2 = \ln(1 + C^2),$$

$$\mu = \ln M - \sigma^2/2.$$

From the normal distribution theory (Chapter 3) it follows that the logarithm of offspring's genotypic value conditional on parental genotypic value is normally distributed with mean and variance

$$E(\ln O|G) = M + h_u^2(\ln G - \mu)/2,$$

$$V(\ln O|G) = (1 - h_u^4/4)\sigma^2,$$

where h_u^2 is the proportion of the additive variance of the total genotypic variance on the underlying scale. We thus have, as with (27), the expressions

$$E(O|G) = M\, G^{h_u^2/2} / \exp\{h_u^2\mu/2 + h_u^4\sigma^2/4\},$$

$$V(O|G) = M^2 G^{h_u^2} (\exp\{(1 - h_u^4/4)\sigma^2\} - 1) / \exp\{h_u^2\mu + h_u^4\sigma^2/2\}$$

for the exact genotypic regression of offspring on one parent, when there is an infinite number of multiplicative loci causing the variation. Clearly, the conditional progeny mean, given parental genotypic value, is a non-linear function of that value, the curve being concave, and the conditional variance is by no means constant, but proportional to the parental value.

Similarly, since the genotypic values 0, G_m and G_f are distributed on log scale as multivariate normal variables, when n is very large, we find from

$$E(\ln O|G_m,G_f) = u + h_u^2(\ln G_m - \mu)/2 + h_u^2(\ln G_f - \mu)/2,$$

and

$$V(\ln O|G_m,G_f) = (1-h_u^4/2)\sigma^2$$

the biparental regression for the corresponding multiplicative model as

$$E(O|G_m,G_f) = M(G_m G_f)^{h_u^2/2} / \exp\{h_u^2\mu - h_u^4\sigma^2/4\},$$

from which, noting that $4G_m G_f = (G_m+G_f)^2 - (G_m-G_f)^2$, a rough approximation for mid-parent regression is

$$E(O|\bar{G}) = M\,(\bar{G} - V_G/2)^{h_u^2} / \exp\{h_u^2\mu - h_u^4\sigma^2/4\}.$$

The regression curve is seen to be concave.

To relate these non-linearities to the earlier ones, we compute the quadratic regressions. For calculating the covariances we require $E(OG)$, $E(OG^2)$ and $E(OG_m G_f)$. Let x and y be independent variables both with a standardised normal distribution. Then a random variable $\rho x+\sqrt{1-\rho^2}\, y$ is also $N(0,1)$ and has a correlation ρ with x. As in (27), we have, noting that $\rho = h_u^2/2$,

$$E(OG) = E(\exp\{x\sigma+\mu\} \exp\{(\rho x+\sqrt{1-\rho^2}\, y)\sigma+\mu\})$$

$$= \exp\{2\mu\} E(\exp\{(1+\rho)x\sigma+\sqrt{1-\rho^2}\, y\sigma\})$$

$$= \exp\{2\mu\} \exp\{(1+\rho)^2\sigma^2/2\} \exp\{(1-\rho^2)\sigma^2/2\}$$

$$= \exp\{2\mu+\sigma^2\} \exp\{\rho\sigma^2\} = M^2 \exp\{h_u^2\sigma^2/2\}$$

Similarly it can be shown that

$$E(OG^2) = M^3\exp\{(1+h_u^2)\sigma^2\}$$

and

$$E(OG_m G_f) = M^3\exp\{h_u^2\sigma^2\}.$$

The amount and type of curvature of the offspring-parent regression has been calculated for various degrees of genotypic variation and proportions of additive variance on the underlying scale. The results are given in Table 7. The regression is always concave, the deviations from linearity being largest for high

Table 7

Non-linearity ($10*S_b$) in offspring-parent regression calculated for various values of coefficient of variation and proportion of additive variation on the underlying scale when the variation is due to an infinite number of multiplicative loci.

		single parent regression				mid-parent regression			
C	h_u^2 =	1.00	0.75	0.50	0.25	1.00	0.75	0.50	0.25
0.05		-0.00	-0.00	-0.01	-0.01	-0.00	-0.00	-0.00	-0.00
0.15		-0.03	-0.04	-0.06	-0.08	-0.00	-0.00	-0.01	-0.03
0.30		-0.10	-0.16	-0.23	-0.31	-0.00	-0.02	-0.07	-0.13

coefficient of variation and small h_u^2's. In general, when there is a very large number of multiplicatively acting loci, the departures of offspring-parent regression from linearity can be considered negligible.

Few Major Loci beside an Infinite Number of Minor Loci

In the additive model we were able to find simple conditions for linearity when loci differed in dominance and allele frequencies. Since all the arguments in that context were based on the symmetrical properties of various third-order moments, we cannot expect to find similar conditions which would lead to linearity under multiplicative gene action, less so if interaction effects are substantial. Instead we pay more attention to the case where there are large differences in the effects between loci, assuming that the variation in a character is caused by few loci with large effects

together with an infinite number of loci with small effects. In terms of gene action between loci, there are four alternatives. The residual variation can be due to either additive or multiplicative loci and the major loci can act either in an additive or multiplicative fashion with the rest. The model with the additive alternative for both sets of loci was studied in the previous chapter. Here we shall first deal with the case where all interactions are of a multiplicative type.

We consider a model with n equal major loci against a background variation accounted by a very large number of loci with small effects. A subscript R is used for parameters describing the residual variation and n for those connected with the major locus effects. Various moments required can be calculated in terms of moments about zero which, under linkage equilibrium, are obtained by multiplying corresponding moments over loci, e.g. $E(OG) = E(OG)_n E(OG)_R$. Hence we have

$$\mu = \mu_n \mu_R,$$

$$V_G = \mu_n^2 V_R + \mu_R^2 V_n + V_n V_R = V_n V_R (C_n^{-2} + C_R^{-2} + 1),$$

$$\mu_{3G} = (V_n V_R)^{3/2} \{(\gamma_{2R} + 3C_R^{-1} + C_R^{-3})\gamma_{1n} + (\gamma_{1R} + 2C_R^{-1})\gamma_{2n} C_n^{-3}\},$$

$$\mu_{4G} = (V_n V_R)^2 \{(\gamma_{2R} + 4\gamma_{1R} C_R^{-1} + 6C_R^{-2} + C_R^{-4})\gamma_{2n} + (\gamma_{2R} + 3\gamma_{1R} C_R^{-1} + 3C_R^{-2})4\gamma_{1n} C_n^{-1} + (\gamma_{2R} + 4\gamma_{1R} C_R^{-1} + 5C_R^{-2})6C_n^{-2} + \gamma_{2R} C_n^{-4}\}$$

$$E(og) = V_n V_R (2h_R^2 C_n^{-1} + 2h_n^2 C_R^{-1} + h_n^2 h_R^2)/4,$$

$$E(og^2) = Cov(O,G^2) - 2\mu E(og)$$

$$= (V_N V_R)^{3/2}\{(b_{3R}+h_R^2C_R^{-1}+C_R^{-1}+C_R^{-3})b_{3n} + (b_{3R}+C_R^{-1})h_n^2C_n^{-1} + (h_R^2C_R^{-1}+b_{3R})C_n^{-1} + b_{3R}C_n^{-3}\},$$

where h_n^2 is written for the ratio $E(og)_n/V_n$ and b_{3n} for $E(og^2)_n/V_n^{3/2}$, h_R^2 and b_{3R} being analogous notations for the residual variation. From the standardised forms of moments, i.e. γ_1, γ_2, $E(og)/\sigma_G^2$, and $E(og^2)/\sigma_G^3$, we see that the properties of the model depend on the relative magnitudes of C_n and C_R. So, for example $E(og)/V_G \sim h_n^2$, if C_R is small compared with C_n.

Under this model, the type and degree of non-linearity in offspring-parent regression has been calculated for two coefficients of residual variation (C_R = 0.05, and 0.15) with various proportions of underlying additive variance. Apart from an infinite number of loci there is assumed to be one major locus segregating, the genetic model for it being the same as in the case of n equal loci with a chosen so that $\ln(1+a)/\sigma_R = 5.0$ on the underlying scale. The results, shown in Table 8, are in good agreement with the corresponding additive model when the allele with an increasing effect is recessive. However, the regression is linear when the plus allele is rare and dominant, although most of the residual variation is additive, and also when the plus allele is very common and dominant and the residual variation is mostly of an additive kind. On the whole, compared with the corresponding additive model multiplicative interation is seen to shift non-linearity towards negative curvature, i.e. to reduce convexity and increase concavity.

Table 8

Non-linearity ($10*S_b$) in offspring-parent regression when there is one major locus acting in a background variation due to a large number of loci with small effects. Loci are interacting multiplicatively. Various gene frequencies, degrees of dominance, coefficients of the residual variation and proportions of underlying additive variance in the residual variation have been considered.

C =0.05 (1+2a=1.568)

			single parent			mid-parent	
q	d	h^2_u = 1.0	0.5	0.1	1.0	0.5	0.1
0.10	-1.00	-4.94	-2.18	0.59	-3.91	-1.48	1.17
	-0.90	-3.68	-0.56	0.03	-2.70	-0.90	0.30
	-0.50	-0.61	-0.19	-0.00	-0.23	-0.02	0.05
	0.00	-0.00	0.04	0.18	-0.00	0.02	0.09
	0.50	0.00	0.07	0.18	-0.04	-0.02	-0.00
	0.90	-0.00	0.01	0.05	-0.16	-0.13	-0.11
	1.00	-0.01	0.00	0.03	-0.19	-0.17	-0.15
0.50	-1.00	-0.03	-0.00	0.00	1.11	1.26	1.40
	-0.90	-0.28	-0.08	-0.12	0.69	0.80	0.91
	-0.50	-0.66	-0.63	-0.60	0.03	0.04	0.05
	0.00	-0.00	-0.01	-0.02	-0.00	-0.00	-0.01
	0.50	0.48	0.25	0.11	-0.02	-0.09	-0.18
	0.90	0.20	0.01	-0.07	-0.52	-0.89	-1.29
	1.00	0.04	-0.05	-0.35	-0.82	-1.31	-1.83
0.90	-1.00	-0.00	-0.17	-0.56	0.17	0.05	0.01
	-0.90	-0.00	-0.20	-0.64	0.03	0.03	0.00
	-0.50	-0.02	-0.34	-1.01	0.03	-0.01	-0.10
	0.00	-0.00	-0.28	-1.09	0.00	-0.13	-0.54
	0.50	0.30	0.00	-0.51	0.13	-0.01	-0.73
	0.90	1.81	0.69	-0.70	1.36	0.41	-1.11
	1.00	2.38	1.18	-0.86	1.88	0.83	-1.17

Table 8 (cont.)

C =0.15 (1+2a=3.216)

			single parent			mid-parent	
q	d	h^2_u = 1.0	0.5	0.1	1.0	0.5	0.1
0.10	-1.00	-6.01	-2.49	0.21	-4.62	-1.57	0.72
	-0.90	-4.59	-2.24	-0.05	-3.24	-1.27	0.00
	-0.50	-0.79	-0.47	-0.23	-0.29	-0.13	-0.03
	0.00	-0.03	-0.01	-0.00	-0.00	-0.00	-0.00
	0.50	-0.01	-0.01	-0.01	-0.05	-0.08	-0.10
	0.90	-0.04	-0.07	-0.10	-0.17	-0.24	-0.30
	1.00	-0.05	-0.09	-0.13	-0.21	-0.28	-0.35
0.50	-1.00	-0.05	-0.12	-0.20	1.05	0.95	0.88
	-0.90	-0.28	-0.42	-0.57	0.63	0.55	0.49
	-0.50	-0.66	-0.87	-0.07	0.02	0.00	-0.00
	0.00	-0.02	-0.11	-0.24	-0.00	-0.03	-0.10
	0.50	0.21	0.01	-0.07	-0.01	-0.16	-0.47
	0.90	0.11	-0.07	-0.69	-0.23	-0.84	-1.79
	1.00	0.04	-0.19	-1.13	-0.36	-1.15	-2.34
0.90	-1.00	-0.01	-0.48	-1.62	0.13	-0.00	-0.14
	-0.90	-0.00	-0.54	-1.78	0.10	-0.01	-0.21
	-0.50	-0.04	-0.72	-2.41	0.02	-0.10	-0.62
	0.00	-0.02	-0.61	-2.67	0.00	-0.25	-1.36
	0.50	0.05	-0.08	-1.86	0.05	-0.07	-1.60
	0.90	0.47	0.12	-1.19	0.40	0.08	-1.16
	1.00	0.62	0.26	-0.76	0.52	0.19	-0.71

We have seen that when there is a very large number of loci with small but similar effects, the offspring-parent regression is linear (or almost linear) and the genotypic distribution symmetrical (or almost symmetrical) despite the interaction components between loci. Hence we conclude that in the models comprising few major loci acting together with a large number of minor loci the main issue with respect to non-linearity is the interaction between the major loci and the genetic background due to loci with infinitesimally small effects.

Summary

i. A quadratic regression has been used to assess the degree of non-linearity in the regression between the genotypic values of offspring and parent for a metric character determined by multiplicative loci with dominance. Coefficient of variation is used as a basis of choosing a reasonable range of values for allelic effects. Only positive interaction between loci has been considered.

ii. Moments required for computing the regression are derived using corresponding expressions for the additive model. The symmetry of absolute values of moments, and consequently of a quadratic coefficient, is distorted by interaction components, which, however, make relatively small contributions to the covariances between offspring and parent.

iii. When the coefficient of genotypic variation is high and

attributed to a small number of loci each with the same effect, gene frequency and dominance, the amount and type of non-linearity (+ for convex, - for concave, and 0 for linear regression) with respect to frequency and dominance relationship of the plus allele is for the single parent regression as follows:

	low	intermediate	high
completely recessive	∿ 0	∿ 0	∿ 0
partially recessive	-	-	∿ 0
additive	∿ 0	∿ 0	∿ 0
partially dominant	∿ 0	+	++
completely dominant	∿ 0	∿ 0	∿ 0

and for the mid-parent regression as follows:

	low	intermediate	high
completely recessive	∿ 0	∿ 0	∿ 0
partially recessive	-	∿ 0	∿ 0
additive	∿ 0	∿ 0	∿ 0
partially dominant	∿ 0	∿ 0	+
completely dominant	∿ 0	-	-

iv. In relation to the number of alleles largest departures from linearity occur when there are only two alleles segregating per locus, non-linearity decreasing when the number of alleles increases. When the number of loci is increased non-linearity decreases, the rate being roughly proportional to 1 / number of loci. When the number is infinite, the genotypic distribution is lognormal and regression concave though almost linear.

v. When there is a locus with a large effect interacting multiplicatively with a background variation due to an infinite number of loci, largest deviations from linearity (always concavity) are expected from a rare recessive allele with an increasing effect if most of the residual variation is additive, or by a very common recessive allele with an increasing effect if most of the residual variation is of a non-additive kind.

5. PHENOTYPIC REGRESSION

We have so far discussed the curvilinearity of offspring-parent regression only in terms of genetic causes. We now move on allowing for environmental deviations and study the non-linearity of regression between phenotypic values in respect to various models of environmental variation. Environmental variance reflects a great variety of causes depending on the character and the organism studied. There are basically two kinds of environmental variation. One which can be traced back into recognisable causes, such as errors of measurement, maternal effects or feeding level, and which therefore can be eliminated by experimental design and another which is generally referred to as intangible variation, that is to say, variability in the internal circumstances of an organism. By definition the mean of environmental deviations for a particular genotype is zero, and it is commonplace in quantitative genetics to assume that the distribution of environmental deviations will be the same for all genotypes. In the following we consider also situations where the environmental distribution, in particular the variance, is dependent on the genotypic value, and vice versa.

Genotypic and Environmental Distribution Independent

We shall first discuss the following problem: if the genotypic regression between offspring and parent is linear, under what circumstances will the regression continue to be linear, when the genotypic values are subject to an independent additive error, i.e. will the corresponding phenotypic regression remain linear. This is

a well-known problem in statistics carrying several different names, such as effect of observational errors on regression analysis, regression with both variables subject to error, regression of true value on observed value. In general the regression will no longer be linear. Only under certain, quite special conditions is linearity going to be unimpaired. Before presenting the theorem on the exact conditions, we introduce some functions which are used in describing statistical distributions.

Let $f(x)$ be the density function of a continuous random variable x (using different notation we could, in principle, do the following for a discrete variable, as well). The expected value of $\exp\{tx\}$ is a function of t given by

$$M(t) = \int_0^\infty \exp\{tx\}\, f(x)\, dx. \tag{28}$$

When this integral exists for a certain range of values of t, we may expand the exponential and integrate term by term, thus obtaining the formula

$$M(t) = 1 + \mu_1' t + \mu_2' t^2/2! + \mu_3' t^3/3! + \cdots,$$

where μ_r' is the moment of order r about zero. The function $M(t)$ is accordingly called a moment-generating function (m.g.f.) of the distribution about the value $x = 0$. Similarly, the m.g.f. about the value $x = a$ is defined as the expected value of $\exp\{t(x-a)\}$. If the logarithm of the m.g.f. can be expanded as a convergent series in powers of t, namely

$$\underline{K}(t) = \ln M(t)$$

$$= K_1 t + K_2 t^2/2! + K_3 t^3/3! + \ldots ,$$

the coefficients, $K_r (r = 1, 2, 3, \ldots)$, are called the cumulants of the distribution and $\underline{K}(t)$ is the cumulant generating function (c.g.f.). It can be shown that the first few cumulants can be given in terms of moments about the mean by the formulae

$$K_1 = \mu_1$$
$$K_2 = \mu_2$$
$$K_3 = \mu_3$$
$$K_4 = \mu_4 - 3\mu_2^2 .$$

Also, the mean and the moments about any value of x can be found by calculating the c.g.f. with respect to any convenient origin. For many density functions the integral in (28) does not exist for real values of t, therefore a more useful auxiliary function, the expected value of exp{itx} (where i is the complex operator), is used. This is called the characteristic function. The results can be readily generalized to the multivariable case.

The errors in the dependent variable do not affect the slope of the regression line, however, they will increase the standard error of the coefficient, but this can be removed by increasing the number of observations for a given value of the independent variable. It is obvious that the attenuation effect of errors in the latter is to diminish the slope. Lindley (1947) has shown that this is not the whole story, for even if the true (or genotypic) regression is

linear, it does not follow that the regression between variables subject to error, say, the observed (or phenotypic) regression is still linear. The problem is not at all trivial, in fact, apart from the case of normal distributions it is not immediately obvious that it would ever be possible for a given joint distribution of genotypic values to imagine a suitable distribution of environmental deviations so that the phenotypic regression between offspring and parent is linear. Lindley has proved that the necessary and sufficient condition under which the observed regression will be linear if that between true values is linear, is that the c.g.f. of the true independent variable is a multiple of the c.g.f. of its error. More precisely, in terms of, $\underline{K}_x$, the c.g.f. of the independent variable and, $\underline{K}_E$, that of the error, we must have

$$\beta \, \underline{K}_E = (\beta - b) \, \underline{K}_x , \tag{29}$$

where β and b are the slopes of the true and observed regressions, respectively. Writing $\underline{K}_G$ for the c.g.f. of the genotypic distribution, we find, given the linearity of offspring on parental genotypic value, that the corresponding phenotypic regression is linear only if we have for the c.g.f. of the environmental distribution, as

$$\underline{K}_E = (h_A^2/2 - h^2/2) \, \underline{K}_G / (h^2/2)$$

$$= (1-H^2) \, \underline{K}_G / H^2,$$

where H^2 is the ratio of genotypic to phenotypic variance (i.e. heritability in broad sense, or degree of genetic determination), $h^2 = V_A/V_P$, and $h_A^2 = V_A/V_G$.

Given a normal genotypic distribution, that is, $\underline{K}_G(t) = ut - V_G t^2/2$, we must have $\underline{K}_E(t) = -V_E t^2/2$, in other words for normal distribution only a normal distribution will preserve the linearity of regression (cf. Curnow, 1960). A weaker version of (29) can be written in terms of cumulants, namely for any r

$$\beta K_{rE} = (\beta - b) K_{rx}. \tag{30}$$

As in the context of genotypic regression we approximate the non-linear phenotypic regression by the mean-square quadratic regression of offspring on parental phenotype, P. Since $Cov(O,P) = Cov(O,G)$ and $Cov(O,P^2) = Cov(O,G^2)$, we have, as in (13), for the quadratic coefficient, as

$$b_2 = (V_P E(og^2) - \mu_{3P} E(og))/(\mu_{4P} - V_P^2)V_P - \mu_{3P}^3,$$

where V_P, μ_{3P}, and μ_{4P} are the variance, third, and fourth moment of the phenotypic distribution, respectively. In a standardised form we have

$$b_2\sigma_P = (E(og^2)/\sigma_P^3 - \gamma_{1P}h^2/2)/(\gamma_{2P} + 2 - \gamma_{1P}^2). \tag{31}$$

Since the genotypic values and environmental deviations are distributed independently of each other, we find

$$\gamma_{1P} = H^3\gamma_{1G} + (1-H^2)^{3/2}\gamma_{1E}$$

$$\gamma_{2P} = H^4\gamma_{2G} + (1-H^2)^2\gamma_{2E} \tag{32}$$

and

$$b_2\sigma_P = \frac{H^3(E(og^2)/\sigma_G^3 - h^2\gamma_{1G}/2) - (1-H^2)^{3/2}\, h^2\gamma_{1E}/2}{H^4\gamma_{2G} + (1-H^2)^2\gamma_{2E} + 2 - (H^3\gamma_{1G} + (1-h^2)^{3/2}\gamma_{1E})^2} \qquad (33)$$

For the degree of non-linearity we find from (31) as (15) that

$$R^2/r^2 - 1 = (H^3E(og^2)/\sigma_G^3 - \gamma_{1P}\, h^2/2)^2/(\gamma_{2P} + 2 - \gamma_{1P}^2)(h^4/4). \qquad (34)$$

Similarly it is obtained for the mid-parent regression

$$b_{\bar{2}}\sigma_{\bar{P}} = 4H^3(E(og^2) + E(og_m g_f))/\sigma_G^3) - 2h^2\gamma_{1P}/\sqrt{2}(\gamma_{2P} + 4 - \gamma_{1P}^2). \qquad (35)$$

Distribution of Environmental Deviations Normal

From (33), expressions for several special cases can be derived. First we deal with a situation where the distribution of environmental deviations is normal, which is a usual assumption in the theory of quantitative genetics. Furthermore, let us assume that the genotypic regression is linear. Then, from (20), $E(og^2) = \mu_{3G}\, h_A^2/2$, and we have from (31)

$$b_2\sigma_P = h_A^2H^3(1-H^2)\gamma_{1G}/2(H^4\gamma_{2G} + 2 - H^6\gamma_{1G}).$$

In other words, when the environmental deviations are normally distributed, only a symmetric genotypic distribution retains the linearity, as can be deduced from the weaker version of Lindley's theorem (30). When the variation in a character is due to few equal non-epistatic loci, the genotypic regression is linear if there is

no dominance or there is complete dominance between alleles (Chapter 3). In the first instance the genotypic distribution is symmetrical only when the alleles are equally frequent. If $q > 0.5$ the phenotypic regression is curved upwards, and the opposite type of curvature follows when $q < 0.5$. If the dominance is complete ($d = \pm 1$), the genotypic symmetry is reached when the recessive allele has a frequency $1/\sqrt{2}$. The regression is curved upwards when $q < 1-1/\sqrt{2}$, given the recessivity of plus allele, and when $q < 1/\sqrt{2}$., given its dominance.

The genotypic regression was also seen to be linear when amongst loci of equal effects the frequencies and dominance relationship between alleles are interchanged. Since the genotypic distribution is then symmetrical, no quadratic type of non-linearity is obtained, although the exact regression may not be linear. When the number of loci affecting the variation in a character becomes very large, a normal genotypic distribution follows, and by definition the regression is linear, as was stated earlier. These results hold also for mid-parent regression, except the case of complete dominance (or recessivity) when the genotypic regression is non-linear.

When the genotypic regression is non-linear, the attenuation effect of error is either to reduce or reinforce the existing curvilinearity. Assuming still a normal environmental distribution, the quadratic coefficient is, from (33),

$$b_2\sigma_P = H^3(E(o g^2)/\sigma_G^3) - \gamma_{1G} h^2/2)/(H^4\gamma_{2G} + 2 - H^6\gamma_{1G}^2).$$

Since with the values of q and d, at which the genotypic regression has largest departures from linearity, the sign of the quadratic coefficient is opposite to that of γ_{1G} and $E(og^2)$, environmental deviations can be expected to cancel out, or with lower values of H^2 to reverse, the curvilinearity of genotypic regression. For the mid-parent regression we find from (35)

$$b_{\bar{2}}\sigma_{\bar{P}} = 4H^3(E(og^2)+E(og_m g_f))/\sigma_G^3 - 2h^2\gamma_{1G}/\sqrt{(2)}(H^4\gamma_{2G} + 4 - H^6\gamma_{1G}^2).$$

Hence when dominance is partial, what was said about single parent regression holds also with the mid-parental. However, when the genotypic non-linearity is due to complete or almost complete dominance, environmental effects are expected to reinforce existing non-linearity.

The results for the non-linearity of phenotypic regression are summarised in Figure 8 at three levels of environmental variation (H = 0.1, 0.4, and 0.7) for the case, where the genotypic variation is due to four equal loci. Because of symmetry only the lower half of frequency range has been considered. When there is no dominance the measure of non-linearity can be compared with Latter's (1965) measure of asymmetry of response for the same model (his Table 3). Although only qualitative comparisons can be made, estimates can be said to be in good agreement in the sense that more intense the selection is, larger the asymmetry is, as would be expected from a quadratic relationship, and also, more extreme the gene frequencies are, stronger the asymmetry is. Figure 8 shows that when the plus allele is recessive, the regression is positively curved, the largest deviations from linearity following when the allele is rare

The relationship between non-linearity of offspring-parent regression and degree of dominance for different gene frequencies when the genotypic variation attributed to four equal loci with no epistasis makes up (a) 10%, (b) 40%, and (c) 70% of the total phenotypic variation. The environmental deviations are normally distributed.

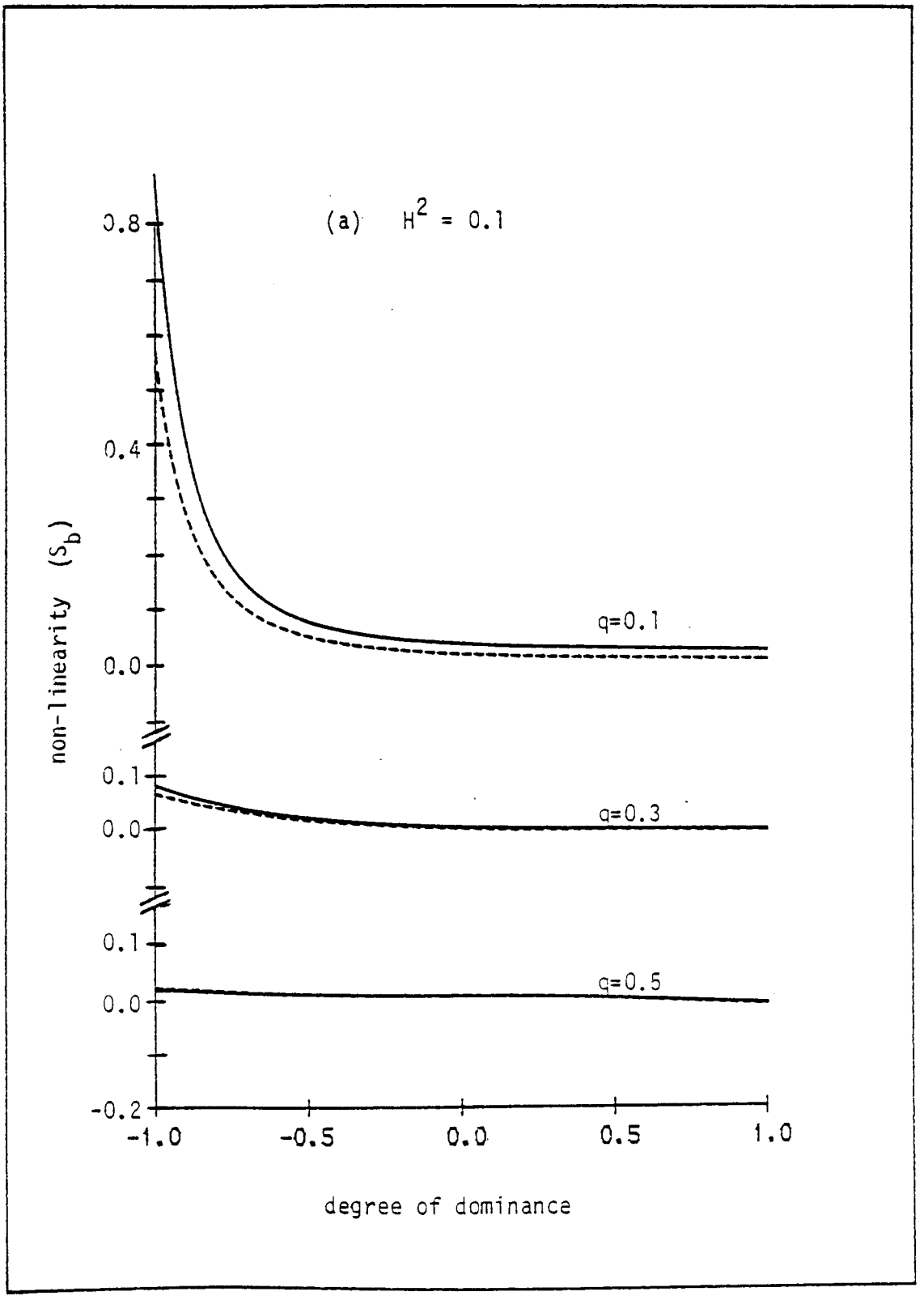


Figure 8 (cont.)

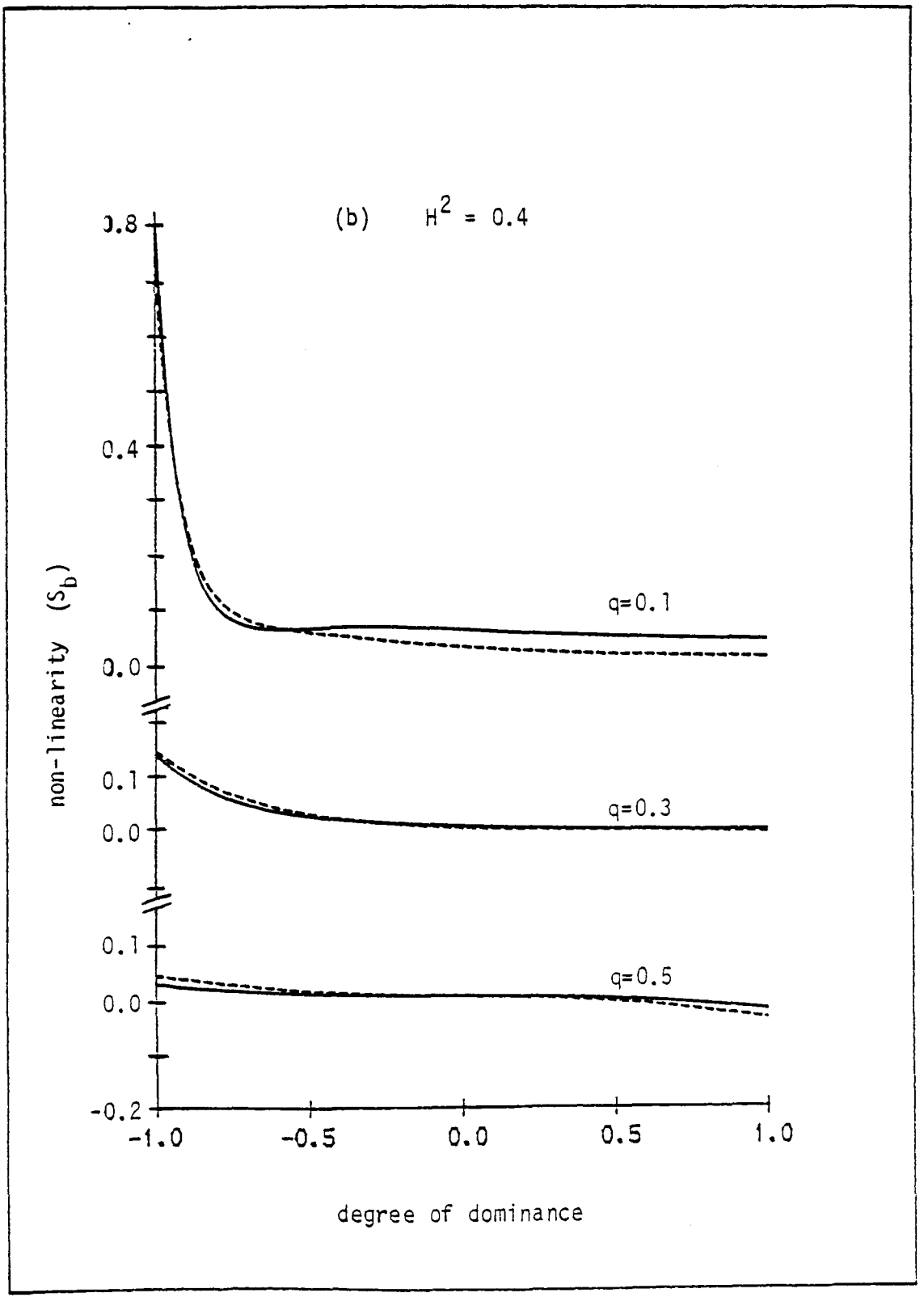
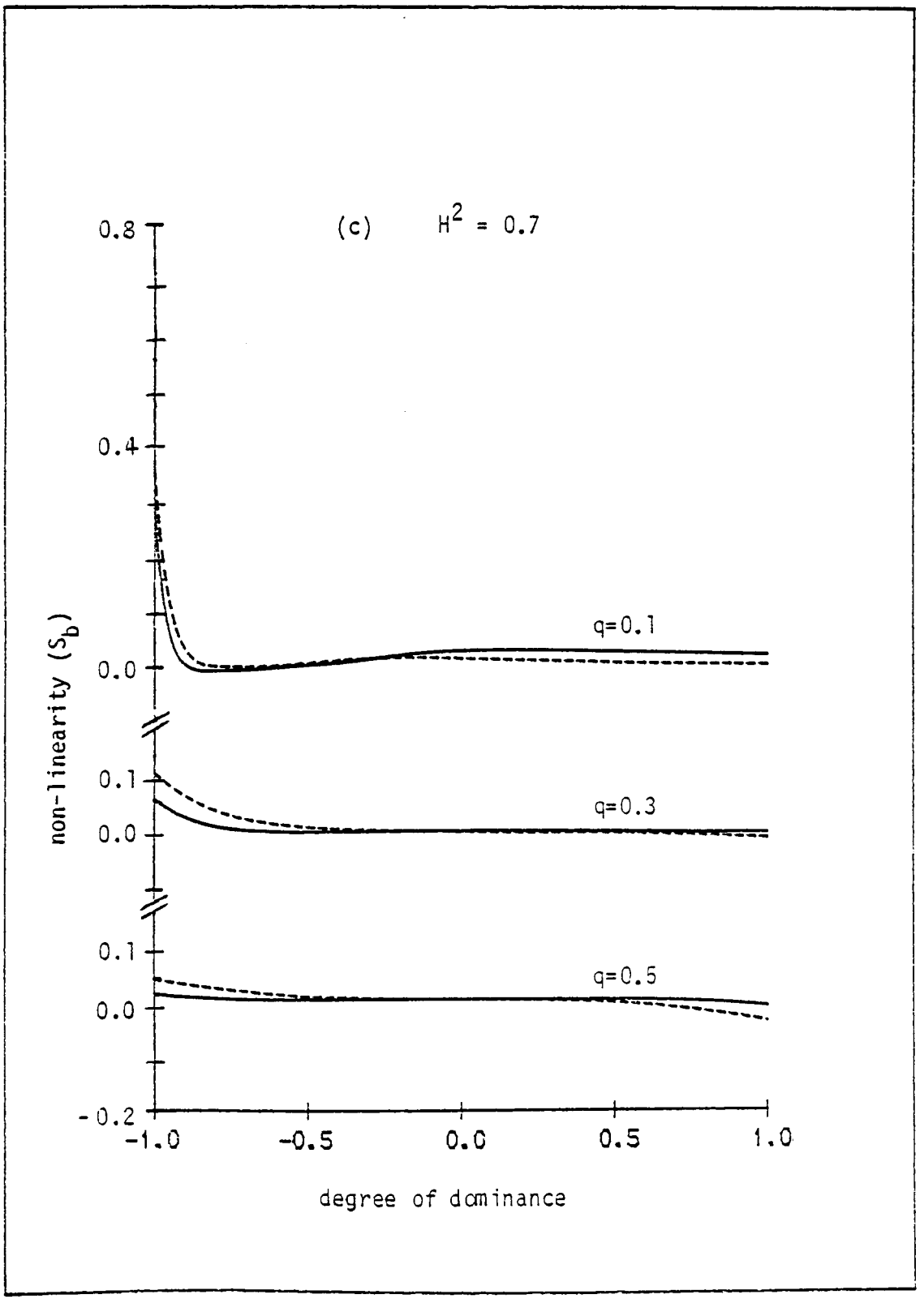


Figure 8 (cont.)



and completely recessive and the proportion of genotypic variance is intermediate or low. Similar conclusions can be drawn for a common dominant allele, the curvature now being downwards. The environmental deviation covers the discrepancies in genotypic non-linearity between single and mid-parent regressions to the extent that even with higher values of H^2 they behave in a similar way. Let us write subscripts 1 and n for parameters describing properties of contributions from a single locus and n equal loci, respectively. Then we have for various expected values of genotypic deviations, as

$$\gamma_{1(n)} = \gamma_{1(1)} / \sqrt{n},$$

$$\gamma_{2(n)} = \gamma_{2(1)} / n,$$

$$E(o g^2)_n / \sigma_n^3 = E(o g^2)_1 / (\sigma_1^3 \sqrt{n}).$$

Hence, from (34), the degree of non-linearity is approximately proportional to 1/n, vanishing when n is very large.

A conspicuous asymmetry, with respect to q and d, was found in the non-linearity of genotypic regression when the loci were interacting in a multiplicative fashion. The departures from linearity due to rare recessives are more noticeable when the allele has a decreasing effect rather than an increasing effect. For the case where the genotypic variation is due to four such loci with equal effects, Figure 9 shows the type and degree of phenotypic non-linearity when the environmental deviations are normally distributed. It is seen that even with substantial interaction

The relationship between non-linearity of offspring-parent regression and degree of dominance for different gene frequencies when the genotypic variation attributed to four equal multiplicative loci (1+2a=1.527) makes up (a) 10%, (b) 40%, and (c) 70% of the total phenotypic variation. The environmental deviations are normally distributed.

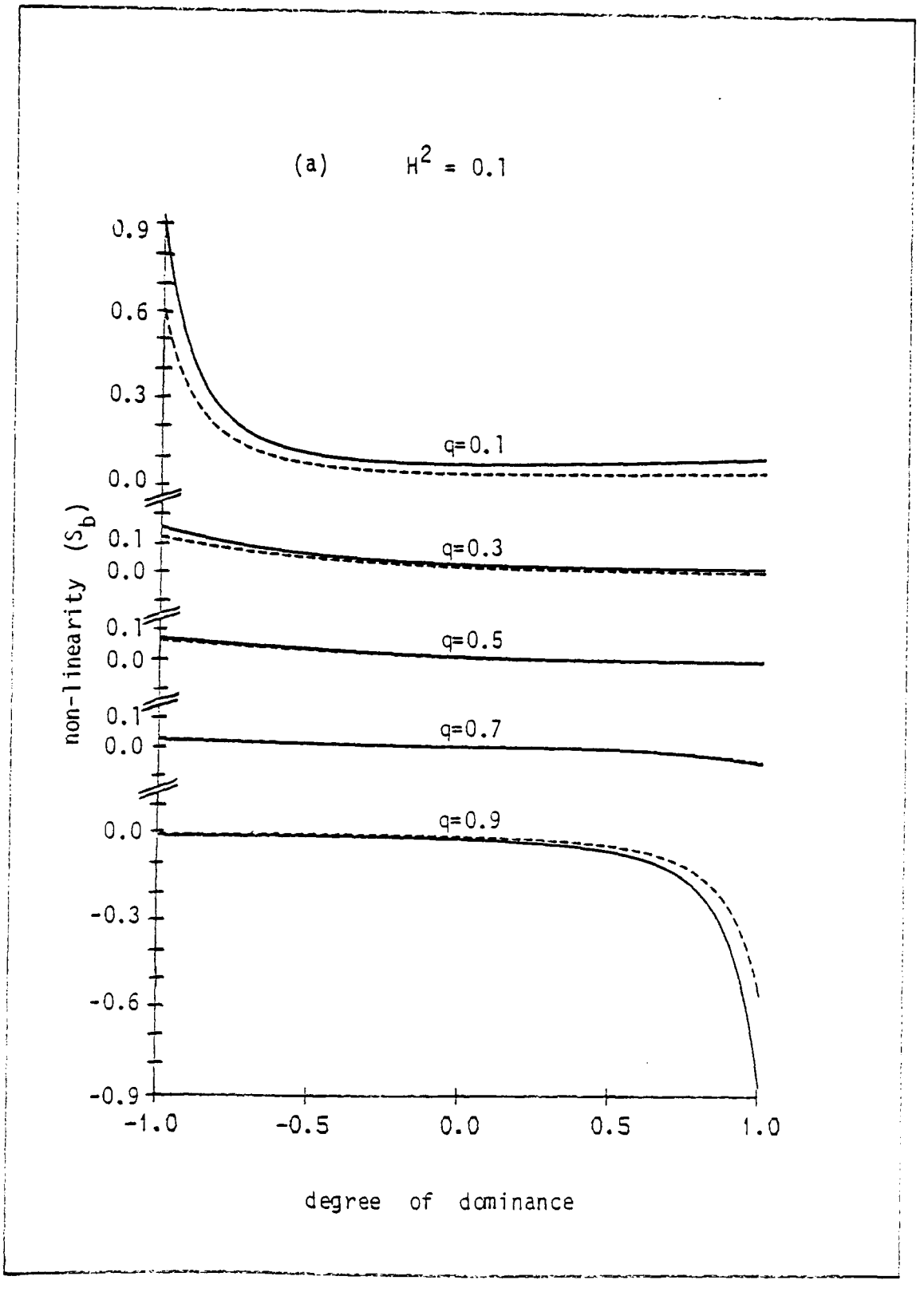


Figure 9 (cont.)

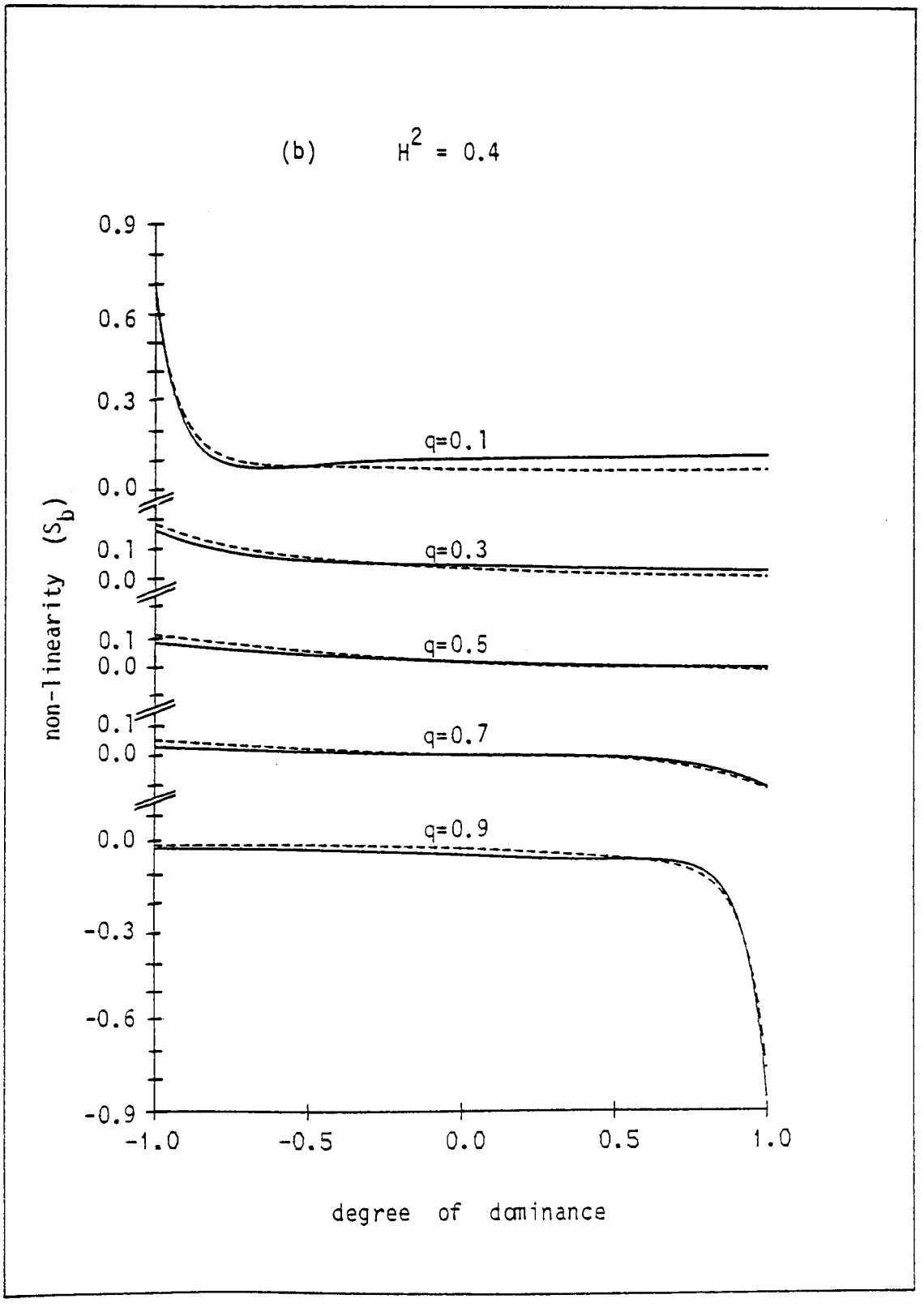
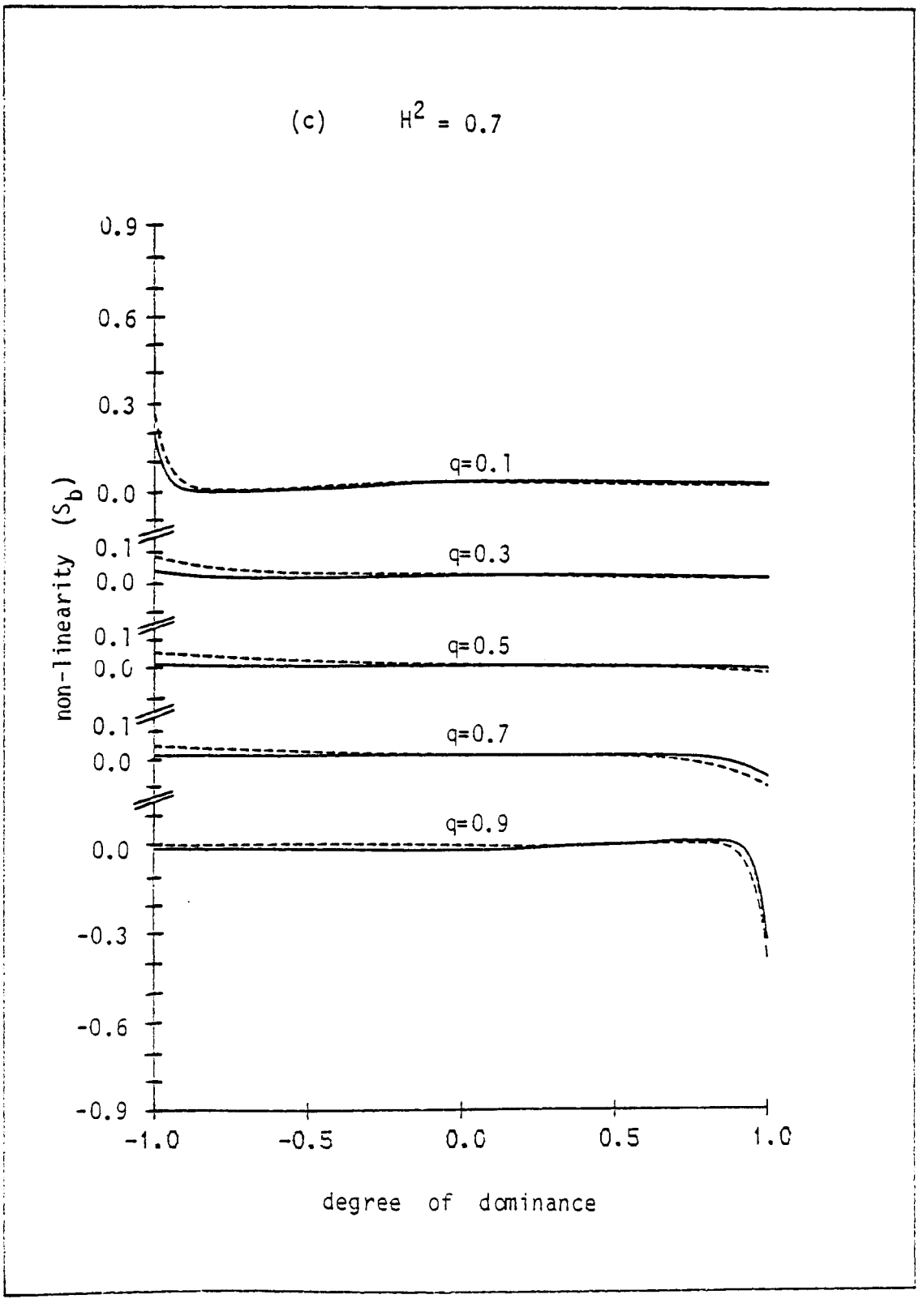


Figure 9 (cont.)



effects non-linearity in the multiplicative model is similar to the one in the additive model when q, d, and H^2 are the same. In other words, largest departures from linearity can be expected when q is very small and d is around -1, and genotypic non-linearity shows up only when H^2 is very large.

Since the third moments in the multiplicative model decline at a rate roughly proportional to 1/n (cf. Table 5), the non-linearity of phenotypic regression can be expected to vanish at a similar rate. However, although the variation is due to a large number of loci, phenotypic regression is going to be slightly convex due to a small positive skewness of the genotypic distribution.

Non-normal Distribution of Environmental Deviations

In this section we study the properties of phenotypic regression when the distribution of environmental deviations is not normal. For example, Clayton (1975) has tentatively concluded that commonly encountered negative skewness in the distribution of egg production in various poultry species is caused by environmental factors, such as disease depressing the performance of affected individuals, rather than by segregation of rare recessive alleles. The skewness of environmental distribution can be attributed also to a scale effect, which in some cases cannot be avoided by transformation, if the genetic and environmental distributions require different scales (e.g. Powers, 1950). The assumption about the independence of genotypic and environmental distribution is maintained.

Given linear genotypic regression the quadratic coefficient

between phenotypes is, from (33),

$$b_2\sigma_P = h^2(1-H^2)(\sqrt{H^2}\gamma_{1G} - \sqrt{(1-H^2)}\gamma_{1E})/2(\gamma_{2P} + 2 - \gamma_{1P}^2). \qquad (36)$$

This is seen to equal to zero when

$$\gamma_{1E} = \sqrt{H^2/(1-H^2)}\gamma_{1G}, \qquad (37)$$

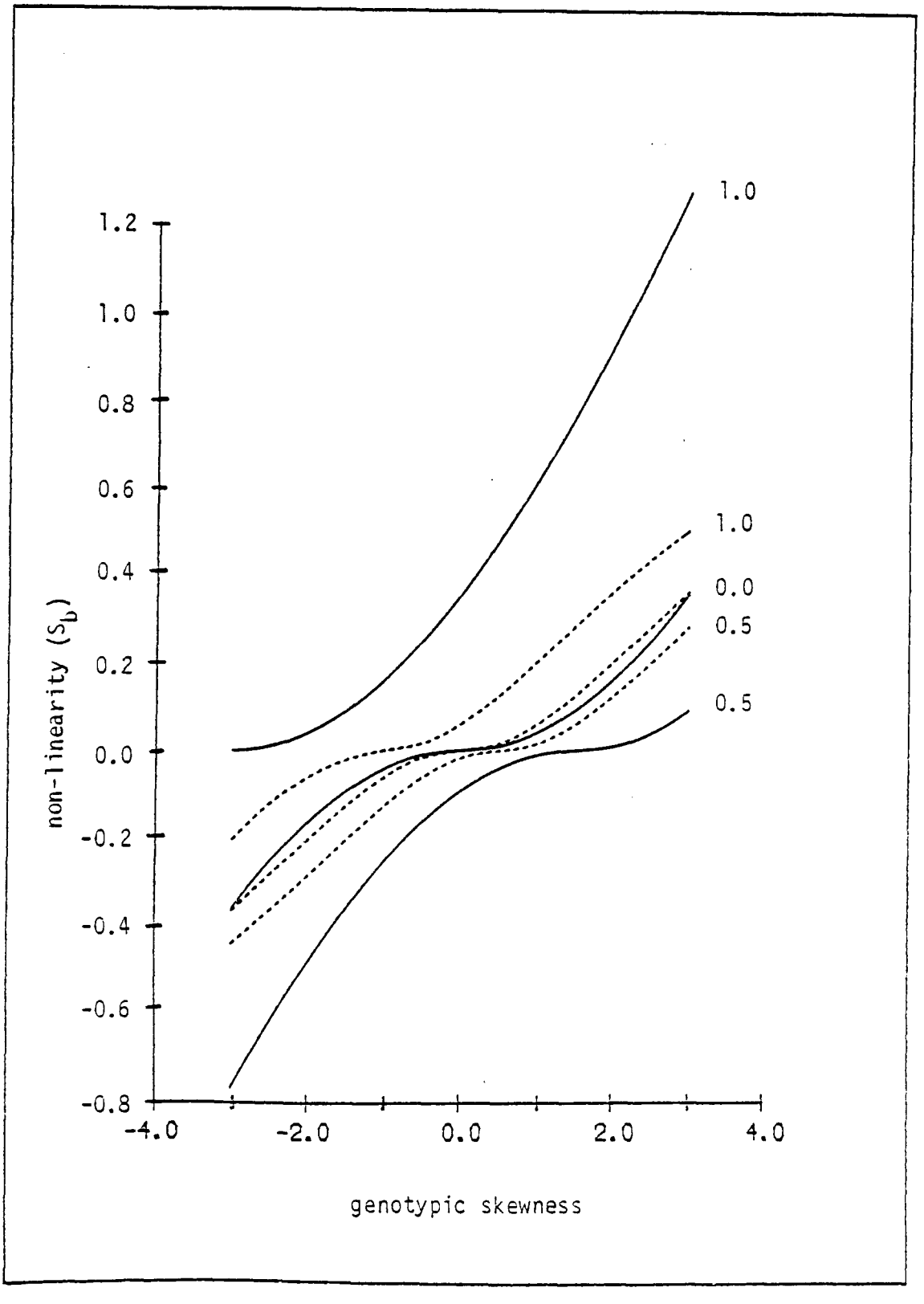
which can be derived also from (30). When, in (37), γ_{1E} is larger than the right-hand side, then the regression is seen to be concave, and when smaller, convexity follows (36), the amount of non-linearity being dependent only of H^2, not of h_A^2. A related result has been empirically discovered by Nishida and Abe (1974), although the narrow range of relatively high H^2's, i.e. from 0.3 to 0.7, made them erroneously conclude that linearity follows whenever the skewnesses of environmental and genotypic distributions are the same. In fact, this is true only when genetic and environmental components contribute equally into the phenotypic variation, since for

$$H^2 = 0.10,\ 0.20,\ 0.30,\ 0.50,\ 0.70,\ 0.90,$$

$$\sqrt{H^2/(1-H^2)} = 0.33,\ 0.50,\ 0.65,\ 1.00,\ 1.20,\ 3.00,$$

respectively. In Figure 10 the degree and type of non-linearity has been plotted against the skewness of genotypic distribution for three different skewnesses, i.e. -2.0, -1.0, and 0.0, of the environmental distribution, at H^2 = 0.1 and 0.5. Binomial distributions with an index 40 and a frequency to produce a required

Figure 10
The relationship between non-linearity in genotype on phenotype regression and genotypic skewness for different values of environmental skewness. Solid lines refer to $H^2=0.1$ and broken lines to $H^2=0.5$.



asymmetry, have been used. Linearity is seen to follow, when (37) is true, e.g. for $H^2 = 0.1$ when $\gamma_{1E} = -1.0$ and $\gamma_{1G} = -3.0$. Deviations from linearity are more noticeable, the larger the difference in the shape of the distributions is. With some H^2's, when environmental and genotypic distributions are skewed in opposite direction, a symmetric phenotypic distribution can follow (cf. 32), the regression being, however, curvilinear, e.g. for $\gamma_{1E} = -1.0$ and $\gamma_{1G} = 3.0$ when $H^2 = 0.1$. This was already pointed out by Nishida and Abe (1974). Also, the smaller the proportion of genotypic variation of the total is, the more substantial the non-linearity is for a given combination of environmental and genotypic distributions. Since the formula (36) is symmetrical in terms of H^2 and skewnesses, environmental distributions skewed only to the left have been considered. For example, the magnitude of curvature (though not the sign) is similar for $\gamma_{1G} = -1.0$, $\gamma_{1E} = -2.0$ and $\gamma_{1G} = 1.0$, $\gamma_{1E} = 2.0$, given the same H^2.

In the most general case we have a non-linear genotypic distribution and an asymmetrical enironmental distribution. Since when the genotypic non-linearity is important the corresponding quadratic coefficient and $E(og^2)$ have different signs, we conclude from (33) that linearity of phenotypic regression is obtained now with environmental skewnesses whose absolute values are slightly smaller than $\sqrt{H^2/(1-H^2)}\ |\gamma_{1G}|$.

Genotypic and Phenotypic Distribution Dependent of Each Other

So far we have discussed cases which are straightforward in the

sense that the distribution of environmental deviations is the same for all genotypes. We have neither considered variation in the environmental levels and consequences of that in genetic variation. In general, genotypic and environmental distributions, or in particular variances, have been assumed to be independent of each other. If the variance of environmental deviations depend on the genotype, then g and e will not be independent, although by definition they are uncorrelated, i.e. $E(ge) = 0$. We also do not want to discard the possibility that the genotypic variance is dependent on the environmental level. In other words, we shall assume $E(ge)$ to be zero, but not necessarily that $E(ge^2)$ and $E(g^2 e)$ are zero. We find for the phenotypic variance, $V_P = V_G + V_E$, where V_G is now the average variance of g over all environments and V_E the average over all genotypes. Assuming linear genotypic regression, the possible non-linearity of offspring-parent regression is then identical to that of the regression g on p. We find expressions for the various moments, as

$$
\begin{aligned}
E(gp) &= V_G, \\
E(gp^2) &= E(g^3) + 2E(g^2e) + E(ge^2), \\
E(p^3) &= E(g^3) + 3E(g^2e) + 3E(ge^2) + E(e^3), \\
V(p^2) &= E(g^4) + 4E(g^3e) + 6E(g^2e^2) + 4E(ge^3) + E(e^4) - V_P^2 .
\end{aligned}
\tag{38}
$$

For the sake of simplicity we deal here only with normal distributions, from which it follows that, e.g. $E(g^3e) = E(ge^3) = 0$. The quadratic coefficient of the regression of genotype on phenotype

is then

$$b_2 = \left[V_P E(gp^2) - \mu_{3P} E(gp)\right] / \left[(V(p^2) - V_P^2)V_P - \mu_{3P}^2\right] .$$

The numerator of this is found to be

$$E(g^2e)(2V_E - V_G) + E(ge^2)(V_E - 2V_G)$$

$$= E(g^2e)(2-3H^2)V_P + E(ge^2)(1-3H^2)V_P . \qquad (39)$$

Let us assume that the environmental variance, or part of it, is dependent of the genotype. Lerner (1954) has argued that inbred lines have an increased environmental variance because homozygotes, compared with heterozygotes are less buffered, or canalized (Waddington, 1942), in their development, and called the phenomenon developmental homeostasis. A recent study in monarch butterflies by Eanes (1978) shows that homozygotes for six enzyme loci, of which at least three are unlinked, generally have larger variances for two forewing characters compared to heterozygotes at the same locus, the means being virtually unchanged. We shall consider here a model where the genes which determine the genetic merit are also responsible of the developmental stability of this character in a way that, for example, individuals with high genotypic values are more sensitive to environmental variation than individuals with low values. Following Robertson (1977), let us suppose that the environmental variance of an individual with a given genotypic value may be written as

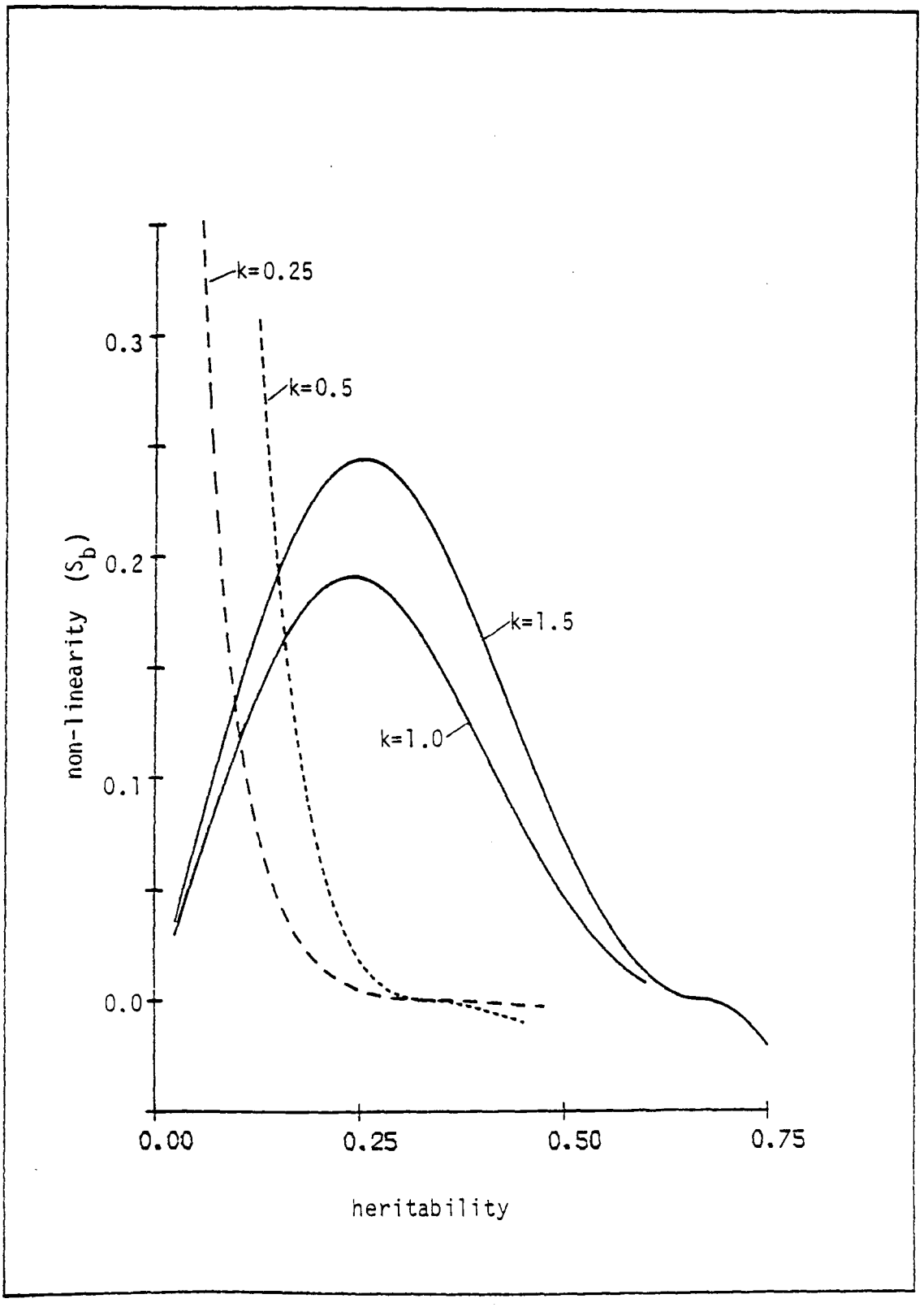
$$E(e^2|g) = e_1^2 + \exp\{kg\},$$

where $e_1 \sim N(0,\sigma_1^2)$ independent of the genotypic value and k is a constant. An exponential function has been chosen not to embody all the complexity of the real world but to feature the biological arguments in a sufficiently simplified form to allow a mathematical analysis of the model. Hence, if we put $V_G = 1$, it is straightforward to show that

$$
\begin{aligned}
V_E &= \sigma_1^2 + \sigma_2^2 \,, \\
E(ge^2) &= k\sigma_2^2 \,, \\
E(g^2e^2) &= (1+k^2)\sigma_2^2 \,,
\end{aligned}
\tag{40}
$$

where $\sigma_2^2 = \exp\{k^2/2\}$. Since e is normal, $\mu_{4E} = 3V_E^2$. From (39), it is seen that the sign of curvature changes at $H^2 = 0.33$. If genotypes with higher values have more developmental 'noise', i.e. $k > 0$, the regression is curved upwards when $H^2 < 0.33$ and downwards otherwise. In terms of selection asymmetry we have higher response upwards when $H^2 < 0.33$, since in selecting downwards amonst selected parents there are also individuals with high genotypic value but large negative environmental deviation. In Figure 11 the degree of non-linearity is shown for values 0.25 and 0.5 of k, consequently we must have $H^2 < 0.469$ and 0.492, respectively (cf. 40). It is seen that the degree of non-linearity is inversely proportional to H^2, being negligible when H^2 is higher. The largest proportion of genetic variance under this model is one half, when $\sigma_1^2 = 0$. When the sign of k is reversed the curvature changes with the amount of non-linearity remaining the same.

Figure 11
The relationship between non-linearity of genotype on phenotype regression and heritability (H^2) when genotypic and environmental distributions are not independent. The solid lines represent models where genetic variation depends on the environmental level, and the broken ones models where environmental variance is not constant over genotypes. For further explanation see the text.



From (40) and (38), the phenotypic distribution is skewed to the right when genotypes with higher values are less stable. In other words starting from normal genotypic distribution we have produced a model contrary to the one with independent distributions under which the quadratic coefficient is negative given the same premises. If we choose a model in which individuals with extreme genotypic values are more susceptible to environmental variation, e.g. $E(e^2|g) = e_1 + kg^2$,we do not find quadratic type of non-linearity, because $E(ge^2) = 0$, instead a cubic regression can be fitted implying lower heritabilities in the tail parts of the phenotypic distribution, which is intuitively obvious.

A second type of dependence between environmental and genotypic distributions can arise if genotypic variation exhibited by individuals depends on the environment they experience. In other words, the heritability may be different at different environmental levels and ranking of individuals on the basis of their genetic merit may not be the same in all environments. To avoid confusion we refer to a type of environment as 'good' ('poor'), if it increases (decreases) the measurement of the trait under study. We assume a special kind of genotype-environment interaction in a sense that although the genotypic values deviate in 'good' and 'poor' conditions to a different extent their ranking is essentially analtered. Usually experiments are so planned that genetic parameters can be estimated directly or indirectly for the context they are going to be used in predicting future performances. However, an insufficient experimental design, due to lack of resources or further knowledge, may cause this kind of effect to be ignored. For example, there is some evidence from dairy cattle,

which suggests that if environments are defined according to the mean level of milk production substantially higher heritabilities for milk yield are obtained in good environments (or at higher performance levels) (Danell, 1982).

Let us divide the environmental deviation into two additive parts, one describing the non-random environmental effects, such as level of nutrition, denoted by e_1, and the other with a constant variance over the environmental range, denoted by e_2. Let us suppose that the genotypic variance for a given environmental level is

$$E(g^2|e_1) = \exp\{ke_1\}.$$

Hence, if we put that $V(e_1) = 1$ and the ratio of $V(e_2)$ to $V(e_1)$ is w, we have, assuming that $E(ge^2) = 0$,

$$V_E = 1 + w,$$

$$V_G = \exp\{k^2/2\},$$

$$E(g^2e) = kV_G,$$

$$E(g^2e^2) = (1+k^2)V_G.$$

From (39) we find that the regression of genotypic value on phenotypic value is linear for $H^2 = 0.67$. In a case where more genetic variation is manifested at a good environment, i.e. $k > 0$, the regression is seen to be curved upwards when $H^2 < 0.67$ and

downwards with higher values of H^2. In Figure 11 the degree of non-linearity is shown at $k = 1.0$ and 1.5, for which $H^2 < 0.622$ and 0.755, respectively. It is seen that the largest deviations from linearity are obtained at intermediate values of H^2. When the sign of k is reversed the sign of quadratic coefficient changes with the magnitude of non-linearity remaining the same.

Whilst in the case of independent distributions the regression is concave when on top of normal genotypic distribution environmental deviations shifted the phenotypic distribution towards positive skewness, we here obtain convexity with the same relationship between skewnesses ($k > 0$).

In the experimental design a lot of attention is usually paid to eliminating unwanted sources of variation by making the environment where individuals are reared as homogeneous as possible and estimating parameters within the same environment. The removal of non-zero $E(ge^2)$ can in some cases be possible through scale transformation. Robertson (1977) has considered a model where $E(ge^2)$ equals $E(g^2e)$. In this case regression is seen to be positively curved when $H^2 > 0.5$ and downwards for values larger than this. If $E(ge^2) > E(g^2e)$, the change in curvature occurs at lower values of H^2.

Summary

i. The effect of environmental deviations on the non-linearity of offspring-parent regression has been studied.

ii. Lindley's general theorem about the effect of an independent additive error on the linearity of regression, namely that linearity is unimpaired only if the cumulant generating function of the error is a multiple of the cumulant generating function of the true independent variable, is reviewed.

iii. If the genotypic regression between offspring and parent is linear and H^2 is the ratio of genotypic to phenotypic variance, the regression on parental phenotype is linear only if the skewness of the environmental distribution is a proportion $\sqrt{H^2/(1-H^2)}$ of that of the genotypic distribution. More the skewnesses depart from this equality or smaller the ratio of genotypic to phenotypic variance is, larger the departures from linearity are. When the environmental skewness is smaller than the value required for linearity, the regression is positively curved and when larger than this value the regression is negatively curved.

iv. The genotypic non-linearity shows up only if H^2 is very large by reinforcing or reducing the curvilinearity caused by difference in skewness between environmental and genotypic distribution. When the genotypic variation is due to a small number of loci each with the same effect, gene frequency and dominance and environmental deviations are normally distributed, the non-linearity of offspring-parent (both single and mid-parent) regression (+ for convex, - for concave, and 0 for linear regression) with respect to frequency and dominance relationship of the plus allele is when H^2 is intermediate or low, as follows:

	rare	intermediate	common
completely recessive	+++	+	-
partially recessive	++	∿ 0	-
additive	+	0	-
partially dominant	+	∿ 0	--
completely dominant	+	-	---

v. The addition of multiplicative interactions between loci does not alter substantially the general picture of phenotypic non-linearity in relation to gene frequency and dominance.

vi. When one or the other of the genotypic extremes has more environmental variance the phenotypic regression is non-linear. If there is more developmental noise amongst individuals with high genotypic values the regression is convex for $H^2 < 0.33$ (smaller the value of H^2 is larger the departures from linearity are) and approximately linear for $H^2 \geq 0.33$.

vii. When the genotypic variation is dependent of the environment individuals experience a non-linear phenotypic regression is found. If the variation is larger in the environment favourable to the measurement a convex regression curve is obtained for $H^2 < 0.67$, with largest departures from linearity being around $H^2 = 0.25$.

6. THE USE OF COLLATERAL RELATIVES IN DETECTING NON-LINEARITY

Introduction

When observations are available on two generations, heritability estimates are obtained by the regression of offspring on parent or response to selection. Linearity of the regression can be checked either by computing the regression only on a selected set of parents or by fitting a non-linear model over the whole phenotypic range. In populations where due to intensive truncation selection only a narrow range of parental phenotypes is used, heritability is usually estimated from the intra-class correlation among sibs making appropriate corrections for the biases caused by selection (e.g. Robertson,1977). The analysis of variance procedure is also used when records in a pedigreed population are available solely on a single generation. Since all records are used in computing sums of squares, checks of linearity equivalent to the offspring-parent ones cannot be made. To overcome this deficiency a method, called 'linear heritability estimation', was proposed by Abplanalp (1961). Individual 'paper' selection is practised and the consequent change is examined at the genetic level in different characters as the mean of their corresponding sib means, weighted according to the proportion of progeny selected. From the selection differential applied, estimates of heritability of the trait under direct selection and of the 'linear' genetic regression of other traits on it can be obtained. The original method by Abplanalp (1961) has been modified by Arthur and Abplanalp (1975). Hill (1978) has examined the bias and sampling variance of the estimator and further

clarified the formulation relating it to the analysis of variance. It is in this latter form we shall discuss the methodology.

Consider a breeding experiment with one-way classification, for example pair matings giving full-sib families or half-sib families of sires each mated to a number of dams each giving rise to one progeny. Let X_{ij} be the trait measurement on individual j in family i, with s families of equal size n. The extension of linear estimator to families of unequal size is straightforward (Hill,1978). The mean of family i and the overall mean are denoted by $X_{i.}$ and $X_{..}$,respectively. As an estimator Abplanalp (1961) considers the regression of family effects, $X_{i.}-X_{..}$, on individual deviations, $X_{ij}-X_{..}$. Since X_{ij} is included in $X_{i.}$, this regression is biassed with it being dependent on the within family variance beside the intra-class correlation. Arthur and Abplanalp (1975) show how to eliminate this bias due to finite family sizes. Their modification is, in essence, to produce family means from which the individual's own measurement is excluded and overall means from which the whole corresponding family is excluded. Let the mean of the sibs of individual j in the family i, itself excluded,be

$$X'_{i.(j)} = (nX_{i.} - X_{ij})/(n - 1)$$

and let

$$X'_{..(i)} = (sX_{..} - X_{i.})/(s - 1)$$

be the mean of individuals unrelated to members of family i. By

applying paper selection among individuals on the basis of their own performance (X_{ij}) we obtain a selection differential, and the mean of the sibs ($X_{i.(j)}$) of these individuals gives a response. Let $q_{ij} = 1$ if the individual is 'selected', and $q_{ij} = 0$ otherwise, with $Q = \Sigma\Sigma_{ij} q_{ij}$ (summation over all individuals) being the total number of selected. Thus the selection differential can be written, as

$$\Sigma_i \Sigma_j q_{ij}(X_{ij} - X'_{..(i)})/Q$$

and the response, as

$$\Sigma_i \Sigma_j q_{ij}(X_{i.(j)} - X'_{..(i)})/Q.$$

The linear estimator proposed by Abplanalp is then the ratio of the latter to the former. In this case it is multiplied by 2 or 4 according to whether family members are full- or half-sibs. The extension of the linear estimator to the hierarchical design is straightforward (Hill, 1978), i.e. we can calculate the ratio of average deviation of either half-sibs or full-sibs from unrelated, or alternatively, the average deviation of full-sibs from half-sibs (dam component), to the average deviation of observations from unrelated individuals.

Hill (1978) has investigated the statistical properties of these estimators when both family effects and within-family deviations are normally distributed. The most noticeable bias is found among extreme ranking individuals causing slight underestimation. It is due to the family structure becoming negligible with increase in

number of families, roughly as 1/s, unless only a few individuals are chosen. A criterion for importance of the influence of family structure seems to be that the selected individuals come from only a few families, such that the small number of families has a substantial effect on the means of top-ranking families. This effect is larger as the intra-class correlation among family members increases, for at high values more of the genetic variation is between families and fewer families are likely to be represented among the chosen individuals (cf. Hill, 1977). Using simulations Hill (1978) has shown that linear estimators computed from an extreme group comprising only part of the total population are making, in terms of sampling variance, an efficient use of the data. His results also show that given fixed total resources, designs which are optimal for conventional estimation of heritability from sibs, are also optimal for estimating linear heritabilities, i.e. n is roughly the inverse of heritability. Although no methods of testing for non-linearity of sib on individual regression in experimental data exist, linear heritabilities can serve as reasonable checks of linearity when data is available only on a single generation. Therefore we would like to investigate, how, to what extent and under what genetic or environmental models, the sib on individual regression depart from linearity and how does it compare in this sense with offspring on parent regression.

Genotypic Regression

Let us consider a random mating population where pedigreed records of a quantitative trait are available on a single

generation. The population size is assumed to be essentially infinite so that biases due to finite family size can be overlooked. We consider three different sib on individual regressions, the regression between either, full-sibs, half-sibs, or full-sibs within half-sibs. To assess the type and amount of non-linearity we compute the quadratic mean-square regression as before (Chapter 2). To allow for comparisons with the offspring-parent regression the various genotypic moments required are written in terms of expectations of powers and products of additive and/or dominance deviations.

Additive Model

Let us suppose that there are n equal non-epistatic loci contributing to the variation in a quantitative trait. Using the notation introduced in the Chapter 3, we write for the genotypic deviation of a randomly chosen individual as $g = \alpha_i + \alpha_k + \delta_{ik}$, and assume that it consequently belongs to a full-sib family with a mean

$$\bar{f} = (\alpha_i+\alpha_j+\alpha_k+\alpha_l)/2 + (\delta_{ik}+\delta_{im}+\delta_{jk}+\delta_{jm})/4$$

or to a half-sib family with a mean $\bar{h} = (\alpha_i+\alpha_j)/2$. From these we find for various covariances, as

$$E(\bar{f}g) = E((\alpha_i+\alpha_k)^2/2 + \delta_{ik}^2/4) = E(A^2)/2 + E(D^2)/4, \qquad (41)$$

$$E(\bar{f}g^2) = E((\alpha_i+\alpha_k)^3/2+(\alpha_i+\alpha_k)^2\delta_{ik}+(\alpha_i+\alpha_k)\delta_{ik}^2/2$$

$$+\delta_{ik}(\alpha_i+\alpha_k)^2/4+(\alpha_i+\alpha_k)\delta_{ik}^2/2+\delta_{ik}^3/4)$$

$$= E(A^3)/2 + 5E(A^2D)/4 + E(AD^2) + E(D^3)/4,$$

whereas the expectations of all other terms equal zero, and

$$E(\bar{h}g) = E(A^2)/4,$$
$$E(\bar{h}g^2) = E(A^3)/4 + E(A^2D)/2 + E(AD^2)/4. \tag{42}$$

It is immediately seen, from (16), (41), and (42) that

$$E(\bar{f}g) = E(og)/2 + V_G/4,$$
$$E(\bar{f}g^2) = E(og^2)/2 + \mu_{3G}/4,$$
$$E(\bar{h}g) = E(og)/2, \tag{43}$$
$$E(\bar{h}g^2) = E(og^2)/2.$$

From (43) we find for the covariances of an individual with the mean of full-sibs within half-sibs, $\bar{f}_h$, as

$$E(\bar{f}_h g) = V_G/4,$$
$$E(\bar{f}_h g^2) = \mu_{3G}/4. \tag{44}$$

Using subcripts F, H, and F|H we have, from (13), (43), and (44), the quadratic coefficients of full-sib, and half-sib, and full-sib within half-sib regressions, respectively, as

$$b_{2F} = (V_G(E(og^2)/2 + \mu_{3G}/4) - \mu_{3G}(E(og)/2 + V_G/4))/T,$$

$$b_{2H} = (V_G E(og^2)/2 - \mu_{3G} E(og)/2)/T, \qquad (45)$$

$$b_{2F/H} = (V_G \mu_{3G}/4 - \mu_{3G} V_G/4)/T = 0,$$

where $T = (\mu_{4G} - V_G^2)V_G - \mu_{3G}^2$. Thus the genotypic regression between full-sibs within half-sibs is always linear irrespective of dominance. Since, from (20) and (45), $2b_{2H}$ equals the quadratic coefficient of single parent regression and the slopes of linear regressions are related in the same way , the type and degree of non-linearity of half-sib and offspring-parent regressions are identical. The full-sib regression has the same quadratic coefficient as the half-sib one, thus exhibiting the same kind of curvature as the single parent regression. However the full-sib estimates on genotypic selection asymmetry are biassed, because linear regression (or intra-class correlation) can greatly overestimate the heritability in the presence of dominance.

Multiplicative model

If the loci affecting a character are interacting multiplicatively the necessary moments can be obtained by means of moments about zero, explained in Chapter 4. Denote the contributions to the genotypic value from the ith locus in a randomly chosen individual and its half-sib by G_i and G_{Hi}, respectively. As previously, the genotypic value is $G = \prod_{i=1}^{n} G_i$. We have for the expected values of various products of contributions from the ith locus, in, say, half-sibs, as

$$E(G_{Hi}G_i) = m_2 + E(A^2)/4,$$

$$E(G_{Hi}G_i^2) = m_3 + mm_2 + mE(A^2)/2 + E(A^2D)/2 + E(AD^2)/4 + E(A^3)/4,$$

where the notations are as in (26). Since multiplicative interaction, or epistasis in general, contributes relatively little into the covariances between relatives and also do not alter drastically the non-linearity caused by dominance, we conclude, almost intuitively, that half-sib regressions are useful estimates of genotypic non-linearity, although there are multiplicative components between loci.

Effects of Linkage

Even though the population is in linkage equilibrium, linkage may show up in certain components of the covariance between relatives. Cockerham (1956), who was the first to investigate the problem, found that the covariances of some relatives are affected, whereas others are not, and that only epistatic components involving sets of linked loci are concerned. His results have since been generalized by Schnell (1963) and Gallais (1974). It is the effect of linkage on the sib on sib regression we shall now consider in more detail, and assume for simplicity that the number of loci involved is effectively infinite.

Provided that the population is in linkage equilibrium, Bulmer (1976) has shown that, in the absence of epistasis, the regression between any pair of relatives is unaffected by linkage. However, the joint distribution of genotypic values in two or more relatives

is no longer multivariate normal, unless the related individuals are monozygotic twins or parent and offspring (Bulmer, 1971). If the loci are multiplicative and unlinked, it follows from the normal distribution theory (cf. Chapter 4) that we have for the half-sibs

$$E(\ln G_H|G) = \mu + h_u^2(\ln G-\mu)/4,$$

$$V(\ln G_H|G) = (1-h_u^4/16)\sigma^2,$$

where μ and σ^2 are the mean and variance, and h_u^2 the proportion of additive variance on the underlying scale. Hence, as with (27),

$$E(G_H|G) = MG^{h_u^2/4}/\exp\{h_u^2\mu/4+h_u^4\sigma^2/32\}, \qquad (46)$$

where $M = \exp\{\mu+\sigma^2/2\}$. Since in the presence of linkage the joint distribution of sibs is no longer normal, this procedure cannot be repeated in deriving the regression function. To illustrate the effect of linkage we shall consider an extreme situation with virtually complete linkage between all pairs of loci and apply a method first suggested by Bulmer (1976). When the linkage is complete the number of pairs of identical genes must be the same at all loci, so that the expected value of G_H given G can be deduced from a mixture of two lognormal distributions: with probability 1/2 there are no genes identical at any locus and G_H will be lognormal with mean M, and with probability 1/2 there is one identical gene at every locus, so that the distribution of G_H given G will be the same as that of offspring given parent, in other words lognormal with mean

Figure 12

4 $E(G_H|G)$ plotted against G for different underlying heritabilities when the variation is due to an infinite number of multiplicative loci (C=0.30). Broken line refers to the case where there is no linkage and the solid one to the case where linkage is complete the latter coincides with the corresponding offspring on one parent regression. The unit length on both axes is one genotypic standard deviation.

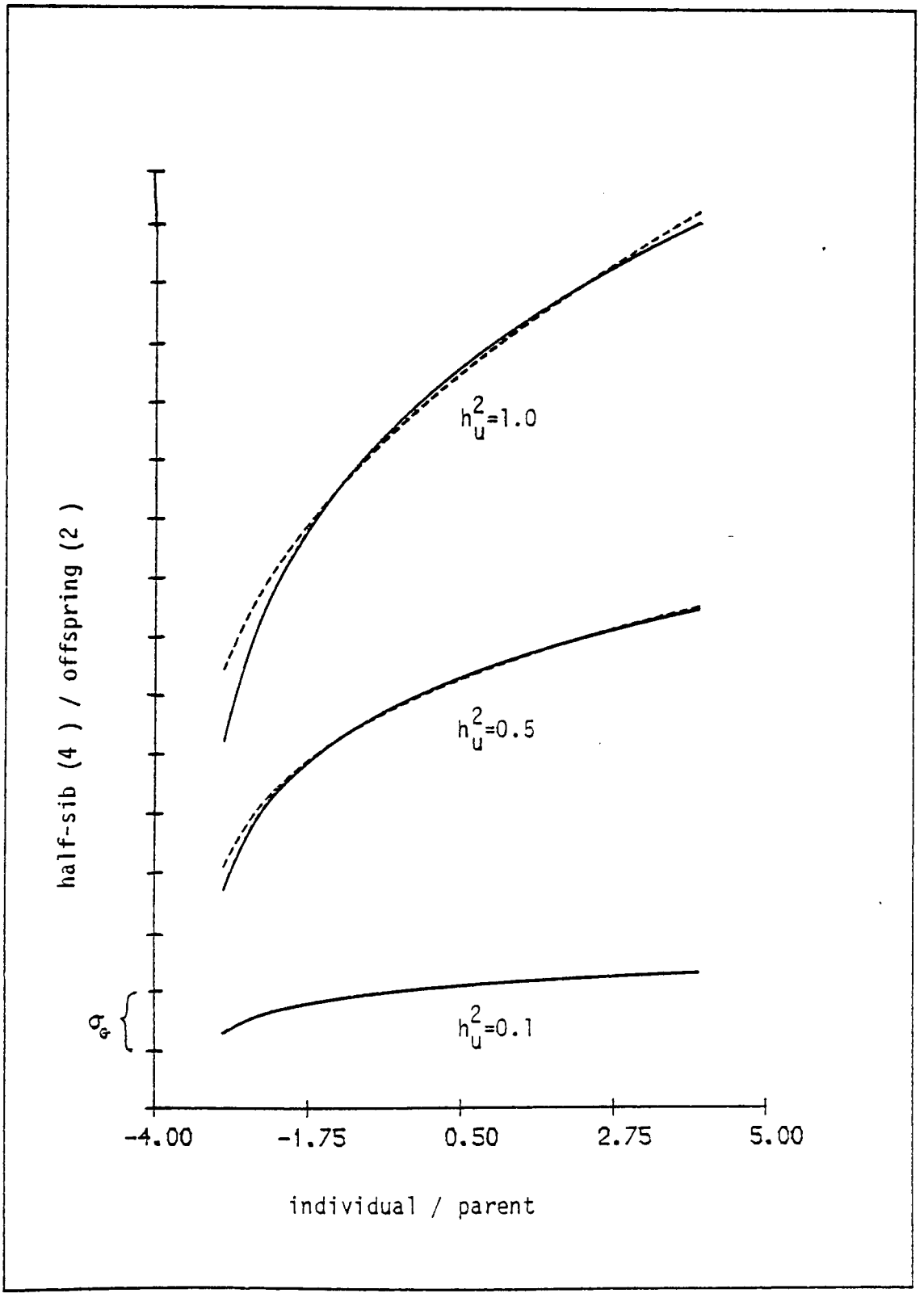
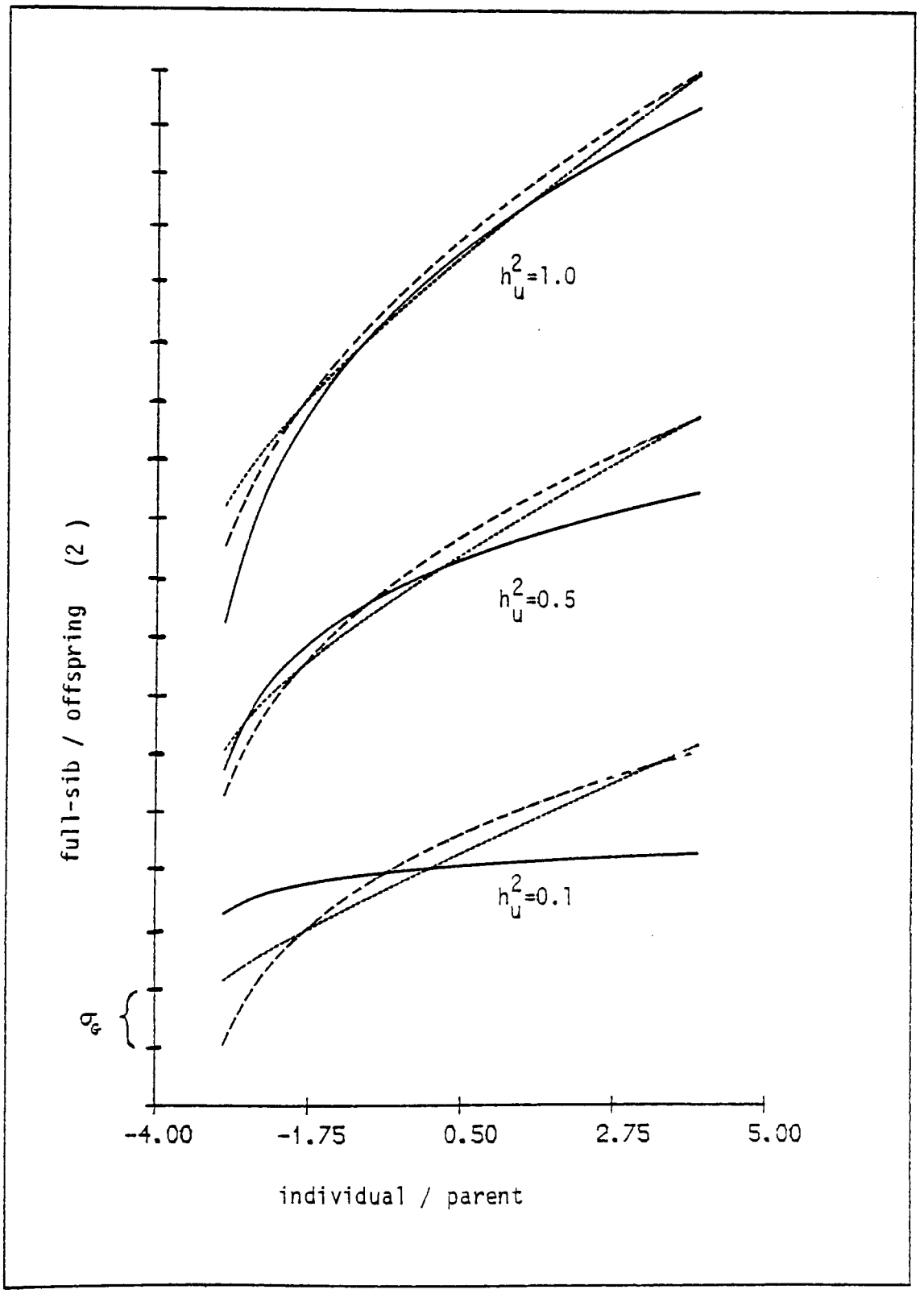


Figure 13

The exact genotypic regression between full-sibs for different underlying heritabilities when the variation is due to an infinite number of multiplicative loci (C=0.30). The broken line refers to the case where there is no linkage and the dotted one to the case where the linkage is complete. The solid line shows the corresponding regression of offspring on single parent. The unit length on both axes equals one genotypic standard deviation.



$$MG^{h_u^2/2}/\exp\{h_u^2\mu/2+h_u^4\sigma^2/4\}.$$

The half-sib regression is therefore

$$E(G_H|G) = M/2 + MG^{h_u^2/2}/2\exp\{h_u^2\mu/2+h_u^4\sigma^2/4\},$$

clearly a different function of G than (46), derived for unlinked loci. The regression functions for offspring and parent, and for half-sibs are shown in Figure 12. The shapes of regression curves are seen to be very similar to each other and the bias - mainly positive - due to linkage negligible.

Similarly, we have for the full-sibs, if loci are unlinked, the expressions

$$E(\ln G_F|G) = \mu + (1+h_u^2)(\ln G-\mu)/4$$

and

$$V(\ln G_F|G) = (1-(1+h_u^2)^2/16)\sigma^2,$$

from which

$$E(G_F|G) = MG^{(1+h_u^2)/4}/4\exp\{(1+h_u^2)\mu/4+(1+h_u^2)^2\sigma^2/32\}$$

When the linkage is complete the expected value of G_F given G can be deduced from a mixture of three distributions: with probability 1/4 there are no genes identical at any locus and G_F will be lognormal with mean M, with probability 1/2 there is one identical gene at every locus, so that the distribution of G_F given G will be the same as that of the offspring given parent, and with

probability 1/4 there are two identical genes at every locus so that $G_F = G$. The full-sib regression is therefore

$$E(G_F|G) = M/4 + MG\,h_u^2/2/2\exp\{h_u^2\mu+h_u^4\sigma^2/4\} + G/4.$$

The regression function for offspring-parent and for full-sibs are shown in Figure 13. Apart from the trivial bias due to dominance the regressions function do not differ much, in other words the positive bias in the tail parts of the distribution can be regarded as negligible.

In the real world linkage is always partial and is likely to be important only in organisms with a small number of chromosomes, such as Drosophila, which in addition has no recombination in males. As Cockerham (1956) has shown the bias in covariances due to linkage is always positive and proportional to the strength of linkage, we therefore conclude by linear interpolation that the effects of partial linkage on half-sib estimates of non-linearity can be regarded as negligibly small.

Phenotypic Regression

When we allow for environmental deviations and compare non-linearity between offspring-parent and sib on sib regression, two main cases can be separated: one when genotypic regression is not linear and the other when it is, that is to say, one where dominance makes a substantial contribution into genetic variation, and the other when most of the variation is additive.

It is straightforward to show that for the quadratic coefficients of phenotypic regressions b_2 of full-sibs = $2b_2$ of half-sibs = b_2 of offspring-parent, and obviously, b_2 of full-sibs within half-sibs equals b_2 of the regression of genotypic on phenotypic value. Hence in the first of our cases only half-sib regressions will give us unbiassed estimates of offspring-parent non-linearity, whereas full-sib and full-sib within half-sib estimates can be used only as a check of the direction - not the magnitude - of curvature, and the latter estimates poorly if heritability is intermediate or high.

If the phenotypic non-linearity is caused by factors other than dominance, then all sib on sib regressions exhibit the same kind of non-linearity as the offspring-parent regression. However, if we have a model where genotypic variation is dependent on the environment individuals experience, the sib on sib regression will grossly underestimate the amount of non-linearity if sibs share the environment as it is often the case with full-sibs.

Summary

i. The use of Abplanalp's linear heritability estimates for checking the asymmetry of selection response under various models of genetic and environmental variation have been studied.

ii. The methodology of linear heritability estimation is reviewed.

iii. Assuming infinitely large families the covariances between various sib means and squared values of individuals are derived in

terms of bivariate moments of additive and dominance deviation.

iv. Non-linearity in half-sib regression is found to be the same as in the regression of offspring on one parent. Due to dominance or common environmental effects full-sib and full-sib within half-sib regression can give grossly biassed estimates of the amount though not the direction of asymmetry.

v. In the presence of multiplicative interaction between loci the difference in genotypic non-linearity between offspring-parent and sib on sib regressions can be regarded as negligible. Linkage is then found to cause slight underestimation of the amount of curvature.

DISCUSSION

After a discovery of substantial convexity in offspring-parent regression for a bristle number in a laboratory population of Drosophila melanogaster studies have been made on the restrictions on the classical prediction equation in quantitative genetics based on conventional heritability estimates. The main interest has been in the circumstances which can lead to non-linearity in offspring-parent regression that is asymmetrical responses in opposite direction in the first generation of selection.

From the algebraic treatment of various models of genetic and environmental variation it is clear that non-linearity can have a variety of causes. The models are of course, rather simplified and it should be asked what relevance these results have to real situations. On the other hand we have omitted some models which inevitably lead into asymmetrical responses in two-way selection. For example, since the phenotypic range of any character is always ultimately bounded by such as a bristle score of zero in Drosophila or a barrier of one egg a day in poultry, non-linearity must in that situation be a rule when the mean is nearing such an extrinsic limit. We have also not considered models which are complicated by differential fitnesses dependent on the value of the character. If the trait is an important component of natural fitness, the regression of offspring on parent is likely to be non-linear in the sense that selection towards increased fitness is giving a smaller response than selection towards decreased fitness. A related phenomenon is the often encountered segregation of lethal alleles in selected lines of Drosophila (e.g. Madalena and Robertson, 1965).

This would again lead into asymmetrical responses, so that response in the direction of previous selection can be close to zero but can be substantial for selection in the opposite direction.

Assuming that genotypic values and environmental deviations are independently distributed the most important contributions to non-linearity are likely to come from rather rare, completely or almost completely recessive alleles segregating at loci with considerable effect on the measurement of the character. Except for the occurrence of dominance in loci determining quantitative traits aspects of the models have been discussed in various sections of the thesis. In discussing the magnitude of contributions from individual loci we concluded that in several studied cases it has been found that a major part of the genetic variation can be attributed to a small number of loci with relatively large effects. From the bottlenecking experiments it can be inferred that the alleles fixed in a selection procedure are usually at intermediate frequencies in the initial population. However, by crossing different selected lines and selecting further, responses beyond the original limits can be obtained (Lopez-Fanjul and Hill, 1973), suggesting that in the base population there are also alleles with fairly large effects but low frequency.

The simplest genetic explanation of the two commonly found and complementary phenomena - heterosis and inbreeding depression - is that there is dominance of gene action at many loci. Assessing from the commonness of these two, dominance can be regarded as an essential feature of genetic models for loci affecting metric characters. An obvious corollary of Fisher's (1930) "Fundamental

Theorem of Natural Selection" is that little additive variance is found in characters such as number of offspring closely connected with fitness, because that kind of genetic variation has been exhausted by natural selection (Robertson, 1955; 1966b). The genetic variation which does exist, will then be non-additive. Hence we conclude that genes which affect important components of fitness are likely to show more dominance, in particular, directional dominance, than genes determining traits like bristles in Drosophila with no direct effect on fitness.

An alternative theoretical approach by Kacser and Burns (1981) leads to similar expectations. They have shown that recessivity of alleles which reduce enzyme activity is an inevitable consequence of the kinetic structure of biochemical networks. It follows from the inherent buffering of such a system that when we are studying the variation at a locus level in a diploid organism the longer and more complex the sequence of enzyme reactions is, the smaller the effect of the second dose of an individual enzyme on output will be compared with the first one, or the less likely the detection of an allele manifecting reduced enzyme activity. Therefore the larger the effect of an allele on anzyme activity the more dominant it is likely to be.

As a practical problem we would like to know how large an experiment is required if we are to have a 90% chance to detect non-linearity using a 5% significance level. Let X be the parental measurement for a trait which is $N(0,\sigma^2)$ and Y the mean of the corresponding family, there being s families each with n progeny. In general, assuming that also X^2 is normally

distributed, it can be inferred from Kendall and Stuart (1973, p. 346) that if a regression is based on s observations and the true coefficient is zero then the sampling variance of the quadratic coefficient is $S.V.(b_2) = \sigma^2_{Y.X^2X} / (s-4)\sigma^2_{X^2.X}$. The variance about regression of the observed family means will, a priori, be made up of two parts (c.f. Latter and Robertson, 1960): the variance of true family means, $(t-R^2)\sigma^2$ (where t is the intra-class correlation of family members) and the variance due to sampling within families. Hence, assuming that s is sufficiently large that terms of order s^{-1} can be ignored relative to one, and noting that $\sigma^2_{X^2.X} = 2\sigma^4$, we have

$$S.V.(b_2) = (t - R^2 + (1-t)/n) / 2s\sigma^2.$$

Assume that males are measured and mated to a random group of females with progeny in half-sib families, in which case $t = h^2/4$ (no environmental covariance of sibs). Under the null hypothesis of linearity $E(b_2) = 0$ (or $R = h^2/2$) and under the alternative hypothesis $E(b_2) = \text{sign}(S_b)*h^2\sqrt{|S_b|}/2\sqrt{2}$ when $\sigma^2 = 1$ (cf. Chapter 2). For one-tail tests with 5% type 1 error and 90% power we require the ratio of difference to its standard error to exceed 2.9, approximately. In Table 9 the number of families required have been calculated for various values of h^2 and S_b.

Table 9

S_b	h^2 = 0.1	0.3	0.5
0.01	62,000	8,400	3,100
0.20	3,100	420	160
0.50	1,200	170	63

Since non-linearities are more likely to occur in connection with small heritabilities, very large experiments are clearly required.

If the true regression is quadratic and more or less homoscedastic, sampling variances could be drastically reduced in an analogous way to conventional offspring-parent regression (e.g. Hill, 1970) by taking observations at three levels, two as near as possible to the extremes of the phenotypic range and one around the mean each containing one third of the observations. In the case of mid-parental regression a further reduction can be obtained by practising assortative mating within each class. These methods can, however, cause considerable bias in the estimates, if the regression is not truly quadratic over the phenotypic range considered or if the regression is substantially affected by assortative mating.

Abplanalp's linear heritability estimates have been found to serve as reasonable checks of selection asymmetry. Especially estimates from half-sibs can be expected to show the same non-linearity as single parent regression. Half-sibs estimates can be used also as checks for the asymmetry of response when selection is practised in both sexes, because difference between single parent and mid-parent regression almost vanishes when H^2 is intermediate or low. Although no methods of testing non-linearity of sib on individual regression in actual data are not yet available, the resemblance to the intra-class correlation would suggest that it might be more efficient than fitting a non-linear offspring-parent regression when heritabilities are low (cf. Robertson, 1959). Further work clearly needs to be done in this direction.

ACKNOWLEDGEMENTS

This Thesis was done under the supervision of Professor Alan Robertson. I am deeply indebted to him for the initial ideas and his guidance during the course of this work. I owe much also to Dr. W.G. Hill, the study of linear heritability estimates stems from his suggestions.

I am deeply grateful to Jim McKay for a lot of good stimulating discussion, and to both, him and his wife, Helen, for invaluable help while still in Edinburgh and after.

I would also like to thank

Professors D.S. Falconer and J.R.S. Fincham for the provision of research facilities,

Daniel Sorensen, for my understanding of the subject I owe a great deal to discussions with him,

Jenny Smith and Marjorie McEwan for computational assistance, and Norma Alexander for experimental assistance,

Tony Shrimpton, Angela Aldridge, and Inge Gerstl for daily encouragement,

Chris Kershaw for reading the manuscript,

Osk. Huttunen Foundation, Finnish Academy and Edinburgh University for financial support.

And what I see, do and write echoes only faintly the world I have learnt from my family and my home village.

REFERENCES

ABPLANALP, H. (1961). Linear heritability estimates. Genet.Res. 2, 439-448.

ANDERSON, V.L. and KEMPTHORNE, O. (1954). A model for the study of quantitative inheriatnce. Genetics 39, 883-898.

ARTHUR, J.A. and ABPLANALP, H. (1975). Linear estimates of heritability and genetic correlation for egg production, body weight, conformation and egg weight in turkeys. Poult.Sci. 54, 11-23.

BEARDSLEY, J.P., BRATTON, R.W. and SALISBURY, G.W. (1950). The curvilinearity of heritability of butterfat production. J.Dairy Sci. 33, 93-97.

BRADFORD, G.E. and VAN VLECK, L.D. (1964). Heritability in relation to selection differential in cattle. Genetics 49, 819-827.

BULMER, M.G. (1971). The effect of selection on genetic variability. Ma.Nat. 105, 201-211.

BULMER, M.G. (1976). Regressions between relatives. Genet.Res. 23, 281-289.

BULMER, M.G. (1980). The Mathematical Theory of Quantitative Genetics. Oxford Univ.Press, Oxford.

CLAYTON, G.A. (1975). Normality of egg production in poultry. Br.Poult.Sci. 16, 431-439.

COCKERHAM, C.C. (1954). An extension of the concept of partioning hereditary variance for analysis of covariance among relatives when epistasis is present. Genetics 39, 859-882.

COCKERHAM, C.C. (1956). Effects of linkage on the covariances between relatives. Genetics 41, 138-141.

COCKERHAM, C.C. and BURROWS, P.M. (1980). Selection limits and strategies. Proc.Natl.Acad.Sci.USA 77, 546-549.

CURNOW, R.N. (1960). The regression of true value on estimated value. Biometrika 47, 457-460.

DANELL, B. (1982). Interaction between genotype and environment in sire evaluation for milk production. Acta Agric. Scand. (in press)

DA SILVA, J.M.P. (1961). Limits of response to selection. Ph.D. Thesis, University of Edinburgh.

EANES, W.R. (1978). Morphological variance and enzyme heterozygosity in the monarch butterfly. Nature,Lond. 276, 263-264.

EWENS, W.J. (1972). The sampling theory of selectively neutral

alleles. Theor.Pop.Biol. 3, 87-112.

FALCONER, D.S. (1973). Replicated selection for body weight in mice. Genet.Res. 22, 291-321.

FALCONER, D.S. (1981). Introduction to Quantitative Genetics (2nd edn). Longman, London.

FISHER, R.A. (1918). The correlation between relatives on the supposition of Mendelian inheritance. Trans.Roy.Soc.Edinburgh 52, 399-433.

FISHER, R.A. (1930). The Genetical Theory of Natural Selection. Clarendon Press, Oxford.

FRANKHAM, R. (1980a). Origin of genetic variation in selection lines. In: Selection Experiments in Laboratory and Domestic Animals.-ed. A.Robertson. Commonwealth Agricultural Bureaux, Slough. Pp. 56-68.

FRANKHAM, R. (1980b). The founder effect and response to artificial selection in Drosophila. In: Selection Experiments in Laboratory and Domestic Animals.-ed. A.Robertson. Commonwealth Agricultural Bureaux, Slough. Pp. 87-90.

FRANKHAM, R. and NURTHEN, R.K. (1981). Forging links between population and quantitative genetics. Theor.Appl.Genet. 59, 231-234.

FRANKHAM, R., BRISCOE, D.A., and NURTHEN, R.K. (1978). Unequal crossing over at the rRNA locus as a source of quantitative genetic variation. Nature,Lond. 272, 80-81.

GALLAIS, A. (1974). Covariance between arbitrary relatives with linkage and epistasis in the case of linkage equilibrium. Biometrics 30, 429-446.

HALDANE, J.B. (1931). A mathematical theory of natural and artificial selection. Part VII. Selection intensity as a function of mortality rate. Proc. Cambridge Phil.Soc. 27, 131-136.

HAMMOND, K. and JAMES, J.W. (1973). Genes of large effect and the shape of the distribution of a quantitative character. Aust.J.biol.Sci. 23, 867-876.

HILL, W.G. (1970). Design of experiments to estimate heritability by the regression of offspring on selected parents. Biometrics 26, 566-571.

HILL, W.G. (1971). Design and efficiency of selection experiments for estimating genetic parameters. Biometrics 27, 293-311.

HILL, W.G. (1977). Order statistics of correlated variables and implications in genetic selection programmes. II. Response to selection.-Biometrics 33, 703-712.

HILL, W.G. (1978). Estimation of heritability by regression using collateral relatives: linear heritability estimation. Genet.Res. 32, 265-274.

HILL, W.G. (1982). Rates of change in quantitative traits from fixation of new mutations. Proc.Natl.Acad.Sci. (in press)

HORNER, T.W., COMSTOCK, R.E. and ROBINSON, H.F. (1955). Non-allelic gene interactions and the interpretation of quantitative genetic data. North Carolina Agr.Exp.Sta.Tech.Bull. 118.

JINKS, J.L. (1979). The biometrical approach to quantitative variation. In: Quantitative Genetic Variation.-ed. J.N. Thompson and J.M. Thoday. Academic Press, London.

KACSER, H. and BURNS, J.A. (1981). The molecular basis of dominance. Genetics 97, 639-666.

KENDALL, M.G. and STUART, A. (1969). The Advanced Theory of Statistics, Vol. 1. Distribution Theory (3rd edn). Griffin, London.

KENDALL, M.G. and STUART, A. (1973). The Advanced Theory of Statistics, Vol. 2. Inference and Relationship (3rd edn). Griffin, London.

KEMPTHORNE, O. (1954). The correlation between relatives in a random mating populations. Proc.Roy.Soc.London B 143,

103-113.

KEMPTHORNE, O. (1960). The biometrical relations between relatives and selection theory. In: Biometrical Genetics.-ed. O. Kempthorne. Pergamon Press, New York. Pp. 12-23.

KEMPTHORNE, O. (1977), Status of quantitative genetic theory. In: Proc.Int.Conf.Quantitative Genetics.-ed. E. Pollak, O. Kempthorne, and T.B. Bailey. Iowa State Univ., Ames, Iowa. Pp. 719-757

KIMURA, M. (1958). On the change of population fitness by natural selection. Heredity 12, 145-167.

KIMURA, M. and CROW, J.F. (1978). Effect of overall phenotypic selection on genetic change at individual loci. Proc.Natl.Acad.Sci.USA 75, 6168-6171.

KOJIMA, K. (1961). Effects of dominance and size of population on response to mass selection. Genet.Res. 2, 177-188.

LANDE, R. (1976). The maintenance of genetic variability by mutation in a polygenic character with linked loci. Genet.Res. 26, 221-235.

LATTER, B.D.H. (1965). The response to artificial selection due to autosomal genes of large effect. I. Changes in gene frequency at an additive locus. Austr.J.Biol.Sci. 18,

585-598.

LATTER, B.D.H. (1969). Models of quantitative genetic variation and computer simulation of selection response. In: Computer Applications in Genetics.-ed. N.E. Morton. University of Hawaii Press, Honolulu. Pp. 55-62.

LATTER, B.D.H. and NOVITSKI, C.E. (1969). Selection in finite population with multiple alleles. I. Limits to directional selection. Genetics 62, 859-876.

LATTER, B.D.H. and ROBERTSON, A. (1960). Experimental design in the estimation of heritability by regression methods. Biometrics 16, 348-353.

LERNER, I.M. (1954). Genetic Homeostasis. Oliver and Boyd, London.

LERNER, I.M. (1958). The Genetic Basis of Selection. Wiley, New York.

LEWONTIN, R.C. (1977). The relevance of molecular biology to plant and animal breeding. In: Proc.Int.Conf.Quantitative Genetics.-ed. E. Pollak, O. Kempthorne, and T.B. Bailey. Iowa State Univ., Ames, Iowa. Pp. 55-62.

LI, C.C. (1976). First Course in Population Genetics. Boxwood, Pacific Grove, California.

LINDLEY, D.V. (1947). Regression lines and the linear functional relationship. J.R.Statist.Soc.Suppl. 9, 218-244.

LOPEZ-.FANJUL, C. and HILL, W.G. (1973). Genetic differences between populations of Drosophila melanogaster for a quantitative trait. I. Laboratory populations. Genet.Res. 22, 51-68.

MADALENA, F.E. and ROBERTSON, A. (1975). Population structure in artificial selection: studies with Drosophila melanogaster. Genet.Res. 24, 51-68.

McMILLAN, I. and ROBERTSON, A. (1974). The power of methods for the detection of major genes affecting quantitative characters. Heredity 32, 349-356.

MATHER, K. and JINKS, J.L. (1971). Biometrical Genetics (2nd edn). Chapman and Hall, London.

MATTHYSSE S., LANGE K. and WAGENER, D.K. (1979). Continuous variation caused by genes with graduated effects. Proc.Natl.Acad.Sci.USA 76, 2862-2865.

MERAT, P. (1968). Distributions de frequences, interpretation du determinisme genetique des caracteres quantitatifs et recheche de 'genes majeurs'. Biometrics 24, 277-293.

MEYER, H.H. and ENFIELD, F.D. (1975). Experimental evidence on limitations of the heritability parameter.

Theor.Appl.Genet. 45, 268-273.

NISHIDA, A. (1972). Some characteristics of parent-offspring regression in body weight of Mus musculus at different ages. Can.J.Genet.Cytol. 14, 293-303.

NISHIDA, A. and ABE, T. (1974). The distribution of genetic and environmental effects and the linearity of heritability. Can.J.Genet.Cytol. 14, 293-303.

PEARSON, K. (1904). Mathematical contributions to the theory of evolution. XII. On a generalized theory of alternative inheritance, with special reference to Mendel's laws. Phil.Trans.Roy.Soc.London A 203, 53-86.

PETERSEN, P.H. (1972). Vekselvirkning mellem arv og miljo i maelkeproduktionen. Lic. afhandl. Den kgl. Veterinaer- og Landbohojskole, Kobenhavn.

PIPER, L.R. and BINDON, B.M. (1982). Proc. World Congress on Sheep and Beef Cattle Breeding. Massey University, N.Z. (in press)

POWERS, L. (1950). Determining scales and the use of transformations in studies on weight per locule of tomato fruit. Biometrics 6, 145-163.

ROBERTSON, A. (1955). Selection in animals: synthesis. Cold Spring Harbor Symp. Quant. Biol. 20, 225-229.

ROBERTSON, A. (1959). Experimental design in the evaluation of genetic parameters. Biometrics 15, 219-226.

ROBERTSON, A. (1960). A theory of limits in artificial selection. Proc.Roy.Soc.London B 153, 234-249.

ROBERTSON, A. (1963). Discussion: Some comments on quantitative genetics theory. In: Statistical Genetics and Plant Breeding.-ed. W.D. Hanson and H.F. Robinson. Natl.Acad.Sci., Natl.Res.Council Pub. 982. Pp. 108-115.

ROBERTSON, A. (1966a). Artificial selection in plants and animals. Proc.Roy.Soc.London B 164, 341-349.

ROBERTSON, A. (1966b). A mathematical model of the culling process in dairy cattle. Anim.Prod. 8, 95-108.

ROBERTSON, A. (1967). The nature of quantitative variation. In: Heritage from Mendel.-ed. R.A. Brink and E.D. Styles. Univ. of Wisconsin Press, Madison. Pp. 265-280.

ROBERTSON, A. (1969). The application of genetics to animal breeding. In: Fifty Years of Genetics.-ed. J.Jinks. Oliver and Boyd, Edinburgh. Pp. 27-36.

ROBERTSON, A. (1970). The state of quantitative genetics in relation to the real world. Proc.19th Annual National Breeders Round Table. Poultry Breeders of America, Kansas City, Missouri. Pp. 1-17.

ROBERTSON, A. (1977). The non-linearity of offspring-parent regression. In: Proc.Int.Conf.Quantitative Genetics.-ed. E. Pollak, O. Kempthorne, and T.B. Bailey. Iowa State Univ., Ames, Iowa. Pp. 297-304.

ROBERTSON, A. (1980). Natural selection and continuous variation. In: Vito Volterra Symposium on Mathematical Models in Biology.-ed. C.Barigozzi. Springer-Verlag, Berlin. Pp. 48-55.

SCHNELL, F.W. (1963). The covariance between relatives in the presence of linkage. In: Statistical Genetics and Plant Breeding.-ed. W.D. Hanson and H.F. Robinson. Natl.Acad.Sci., Natl.Res.Council Pub. 982. Pp. 468-483.

SHRIMPTON, A.E. (1982). The isolation of polygenic factors controlling bristle score in Drosophila melanogaster. Ph.D. Thesis, Edinburgh University.

SILVELA, L. (1980). Genetic changes with generations of artificial selection. Genetics 95, 769-782.

SORENSEN, D.A. (1980). Effect of truncation selection on genetic variability. Ph.D. Thesis, Edinburgh University.

STROBECK, C. and MORGAN, K. (1978). The effect of intragenic recombination on the number of alleles in a finite population. Genetics 88, 829-844.

THODAY, J.M. (1979). Polygene mapping: uses and limitations. In: Quantitative Genetic Variation.-ed. J.N. Thompson,Jr. and J.M. Thoday. Academic Press, London.

THODAY, J.M., GIBSON, J.B., and SPICKETT, S.G. (1964). Regular responses to selection. 2. Recombination and acclerated response. Genet.Res. 5, 1-19.

WADDINGTON, C.H. (1942). Canalisation of development and the inheritance of acquired characters. Nature,Lond. 150, 563-565.

VAN VLECK, L.D. and HART, C.L. (1965). Linearity of heritability of Holstein first-lactation milk production. J.Dairy Sci. 48, 1376-1377.

WATT, W.B. (1972). Intragenic recombination as a source of population genetic variability. Am.Nat. 106, 737-753

WEHRHAHN, A. and ALLARD, R.W. (1965). The detection and measurement of individual genes involved in the inheritance of a quantitative character in wheat. Genetics 51, 109-119.

WRIGHT, A.J. (1979). The use of differential coefficients in the development and interpretation of quantiative genetic models. Heredity 43, 1-8.

WRIGHT, S. (1921). Systems of mating. Genetics 6, 111-178.